Bayes factors for testing equality and inequality constrained hypotheses on variances

Böing-Messing, Florian

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

BAYES FACTORS FOR TESTING EQUALITY AND INEQUALITY CONSTRAINED HYPOTHESES ON VARIANCES

FLORIAN BÖING-MESSING

Bayes Factors for Testing Equality and Inequality Constrained Hypotheses on Variances

Florian Böing-Messing

ISBN: 978-94-6295-743-5
Printed by: ProefschriftMaken, Vianen, the Netherlands
Cover design: Philipp Alings

Bayes Factors for Testing Equality and Inequality Constrained Hypotheses on Variances

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. E.H.L. Aarts, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op

vrijdag 6 oktober 2017 om 14.00 uur

door

Florian Böing-Messing

geboren te Bocholt, Duitsland

Promotores:	prof. dr. J.K. Vermunt
	prof. dr. M.A.L.M. van Assen
Copromotor:	dr. ir. J. Mulder
Promotiecommissie:	prof. dr. J.J.A. Denissen
	prof. dr. ir. J.-P. Fox
	prof. dr. I. Klugkist
	prof. dr. E.M. Wagenmakers

To my parents, Gaby and Georg

Contents

Chapter 1

Introduction

Statistical data analysis commonly focuses on measures of central tendency like means and regression coefficients. Measures such as variances that capture the heterogeneity of observations usually do not receive much attention. In fact, variances are often regarded as nuisance parameters that need to be "eliminated" when making inferences about mean and regression parameters. In this dissertation we argue that variances are more than just nuisance parameters (see also Carroll, 2003): Patterns in variances are frequently encountered in practice, which requires that researchers carefully model and interpret the variability. By disregarding the variability, researchers may overlook important information in the data, which may result in misleading conclusions from the analysis of the data. For example, psychological research has found males to be considerably overrepresented at the lower and upper end of psychological scales measuring cognitive characteristics (e.g. Arden & Plomin, 2006; Borkenau, Hřebíčková, Kuppens, Realo, & Allik, 2013; Feingold, 1992). To understand this finding, it is not sufficient to inspect the means of the groups of males and females. Rather, an inspection of the variances reveals that the overrepresentation of the males in the tails of the distribution is due to males being more variable in their cognitive characteristics than females.

1.1 Motivating Example

There are often reasons to expect certain patterns in variances. For example, Aunola, Leskinen, Lerkkanen, and Nurmi (2004) hypothesized that the variability of students' mathematical performances either increases or decreases across grades. On the one hand, the authors expected that an increase in variability might occur because students with high mathematical potential improve their performances over time more than students with low potential. On the other hand, they reasoned that the variability of mathematical performances might decrease across grades because systematic instruction at school helps students with low mathematical potential catch up, which makes students more homogeneous in their mathematical performances. These two competing expectations can be expressed as inequality constrained hypotheses on the

variances of mathematical performances in $J \geqslant 2$ grades:

$$\begin{aligned} H_1 &: \sigma_1^2 < \cdots < \sigma_J^2 \quad \text{and} \\ H_2 &: \sigma_J^2 < \cdots < \sigma_1^2, \end{aligned} \tag{1.1}$$

where σ_j^2 is the variance of mathematical performances in grade j, for $j = 1, \ldots, J$. Thus, H_1 states an increase in variances across grades, whereas H_2 states a decrease. Two additional competing hypotheses that are conceivable in this example are

$$\begin{aligned} H_0 &: \sigma_1^2 = \cdots = \sigma_J^2 \quad \text{and} \\ H_3 &: \text{not } (H_0 \text{ or } H_1 \text{ or } H_2), \end{aligned} \tag{1.2}$$

where H_0 is the null hypothesis that states equality of variances and H_3 is the complement of H_0, H_1, and H_2. The complement covers all possible hypotheses except H_0, H_1, and H_2 and is often included as a safeguard in case none of H_0, H_1, and H_2 is supported by the data. Note that we do not impose any constraints on the mean parameters of the grades, which is why these parameters are omitted from the formulation of the hypotheses in Equations (1.1) and (1.2). This illustrates that we reverse common statistical practice in this dissertation by focusing on the variances, while treating the means as nuisance parameters.

1.2 The Bayes Factor

In this dissertation we use the Bayes factor to test equality and inequality constrained hypotheses on variances. The Bayes factor is a Bayesian hypothesis testing and model selection criterion that was introduced by Harold Jeffreys in a 1935 article and in his book *Theory of Probability* (1961). For the moment, suppose there are two competing hypotheses H_1 and H_2 under consideration (i.e. it is assumed that either H_1 or H_2 is true). Jeffreys introduced the Bayes factor for testing H_1 against H_2 as the ratio of the posterior to the prior odds for H_1 against H_2:

$$B_{12} = \frac{P(H_1|\boldsymbol{x})}{P(H_2|\boldsymbol{x})} \Big/ \frac{P(H_1)}{P(H_2)}, \tag{1.3}$$

where $\boldsymbol{x}$ are the data, and $P(H_t|\boldsymbol{x})$ and $P(H_t)$ are the posterior and the prior probability of H_t, for $t = 1, 2$. A Bayes factor of $B_{12} > 1$ indicates evidence in favor of H_1 because then the posterior odds for H_1 are greater than the prior odds (i.e. the data increased the odds for H_1). Likewise, a Bayes factor of $B_{12} < 1$ indicates evidence in favor of H_2.

The prior probabilities $P(H_1)$ and $P(H_2) = 1 - P(H_1)$ need to be determined by the researcher before observing the data and reflect to what extent one hypothesis is favored over the other a priori. In case no hypothesis is favored, a researcher may specify equal prior probabilities of $P(H_1) = P(H_2) = 1/2$, resulting in prior odds of $P(H_1)/P(H_2) = 1$. In this case the Bayes factor is equal to the posterior odds. The posterior probabilities of the hypotheses are obtained by updating the prior probabilities with the information from the data using Bayes's theorem:

$$P(H_t|\boldsymbol{x}) = \frac{m_t(\boldsymbol{x})P(H_t)}{m_1(\boldsymbol{x})P(H_1) + m_2(\boldsymbol{x})P(H_2)}, \quad t = 1, 2, \tag{1.4}$$

where $m_t(\boldsymbol{x})$ is the marginal likelihood of the observed data $\boldsymbol{x}$ under H_t. The posterior probabilities quantify how plausible the hypotheses are after observing the data. In Equation (1.4) the marginal likelihoods are obtained by integrating the likelihood with respect to the prior distribution of the model parameters under the two hypotheses:

$$m_t(\boldsymbol{x}) = \int f_t(\boldsymbol{x}|\boldsymbol{\theta}_t)\pi_t(\boldsymbol{\theta}_t)d\boldsymbol{\theta}_t, \quad t = 1, 2, \tag{1.5}$$

where $f_t(\boldsymbol{x}|\boldsymbol{\theta}_t)$ is the likelihood under H_t and $\pi_t(\boldsymbol{\theta}_t)$ is the prior distribution of the model parameters $\boldsymbol{\theta}_t$ under H_t. In this dissertation we use the normal distribution to model the data. The expression in Equation (1.5) can be interpreted as the average likelihood under hypothesis H_t, weighted according to the prior $\pi_t(\boldsymbol{\theta}_t)$. The marginal likelihood quantifies how well a hypothesis was able to predict the data that were actually observed; the better a hypothesis was able to predict the data, the larger the marginal likelihood.

When plugging the expression for the posterior probabilities of the hypotheses in Equation (1.4) into Equation (1.3), the expression for the Bayes factor of H_1 against H_2 simplifies to the ratio of the marginal likelihoods under the two competing hypotheses:

$$B_{12} = \frac{m_1(\boldsymbol{x})}{m_2(\boldsymbol{x})}. \tag{1.6}$$

Note that the prior probabilities of the hypotheses cancel out in this step, which shows that the Bayes factor does not depend the prior probabilities. From the expression in Equation (1.6) it can be seen the Bayes factor can be interpreted as a ratio of weighted average likelihoods: If $B_{12} > 1$ ($B_{12} < 1$), then it is more likely that the data were generated under hypothesis H_1 (H_2). For example, a Bayes factor of $B_{12} = 10$ indicates that it is 10 times more likely that the data originate from H_1 than from H_2. In other words, the evidence in favor of H_1 is 10 times as strong as the evidence in favor of H_2. Likewise, a Bayes factor of $B_{12} = 1/10$ indicates that H_2 is 10 times more likely.

It is straightforward to test $T > 2$ hypotheses simultaneously using the Bayes factor (as in the motivating example in Section 1.1). In such a multiple hypothesis test the Bayes factor of two competing hypotheses H_t and $H_{t'}$, for $t, t' \in \{1, \ldots, T\}$, is still given by the ratio of the marginal likelihoods under the two hypotheses, that is, $B_{tt'} = m_t(\boldsymbol{x})/m_{t'}(\boldsymbol{x})$. The posterior probabilities of the hypotheses can be computed as $P(H_t|\boldsymbol{x}) = m_t(\boldsymbol{x})P(H_t)/\big(\sum_{t'=1}^{T} m_{t'}(\boldsymbol{x})P(H_{t'})\big)$, for $t = 1, \ldots, T$. Here the prior probabilities $P(H_1), \ldots, P(H_T)$ need to sum to 1, which implies that it is assumed that one of the T hypotheses under investigation is the true hypothesis. A common choice when prior information is absent is to set equal prior probabilities $P(H_1) = \cdots = P(H_T) = 1/T$. In a multiple hypothesis test it is useful to inspect the posterior probabilities of the hypotheses to see at a glance which hypothesis receives strongest support from the data.

From Equation (1.5) it can be seen that in order to compute the marginal likelihoods a prior distribution of the model parameters is needed under each hypothesis to be tested. In fact, Bayes factors are sensitive to the exact choice of the priors. It is therefore crucial to specify the priors with care. In case prior information about the magnitude of the variances is available (e.g. from earlier studies), one might consider using this information to specify informative priors. However, often such prior

information is not available or a researcher would like to refrain from using informative priors (e.g. to "let the data speak for themselves"). In Bayesian estimation it is then common to use improper priors that essentially contain no information about the model parameters. In Bayesian hypothesis testing, however, one may not use improper priors because these depend on undefined constants, as a consequence of which the Bayes factor would depend on undefined constants as well. Using vague proper priors with very large variances to represent absence of prior information is not a solution to this problem when testing hypotheses with equality constraints on the variances. The reason is that using vague priors might induce the Jeffreys–Lindley paradox (Jeffreys, 1961; Lindley, 1957) where the Bayes factor always favors the null hypothesis regardless of the data. Hence, the main objective of this dissertation is to develop Bayes factors for testing equality and inequality constrained hypotheses on variances that can be applied when prior information about the magnitude of the variances is absent. In general, the Bayes factors we propose are based on proper priors that contain minimal information, which avoids the problem of undefined constants in the Bayes factors and the Jeffreys–Lindley paradox. In Chapters 2, 3, and 4 we use a minimal amount of the information in the sample data to specify proper priors in an automatic fashion. In Chapter 5 we propose a default prior containing minimal information based on theoretical considerations.

1.3 Outline of the Dissertation

This dissertation is structured as follows. In Chapter 2 we consider the problem of testing (in)equality constrained hypotheses on the variances of two independent populations. We shall be interested in testing the following hypotheses on the two variances: the variances are equal, population 1 has smaller variance than population 2, and population 1 has larger variance than population 2. We consider three different Bayes factors for this multiple hypothesis test: The first is the fractional Bayes factor (FBF) of O'Hagan (1995), which is a general approach to computing Bayes factors when prior information is absent. The FBF is inspired by partial Bayes factors, where proper priors are obtained using a part of the sample data. It is shown that the FBF may not properly incorporate the parsimony of the inequality constrained hypotheses. As an alternative, we propose a balanced Bayes factor (BBF), which is based on identical priors for the two variances. We use a procedure inspired by the FBF to specify the hyperparameters of this balanced prior in an automatic fashion using information from the sample data. Following this, we propose an adjusted fractional Bayes factor (aFBF) in which the marginal likelihood of the FBF is adjusted such that the two possible orderings of the variances are equally likely a priori. Unlike the FBF, both the BBF and the aFBF always incorporate the parsimony of the inequality constrained hypotheses. In a simulation study, the FBF and the BBF provided somewhat stronger evidence in favor of a true equality constrained hypothesis than the aFBF, whereas the aFBF yielded slightly stronger evidence in favor of a true inequality constrained hypothesis. We apply the Bayes factors to empirical data from two studies investigating the variability of intelligence in children and the precision of burn wound assessments.

In Chapter 3 we address the problem of testing equality and inequality constrained hypotheses on the variances of $J \geqslant 2$ independent populations. Hypotheses on the

variances may be formulated using a combination of equality constraints, inequality constraints, and no constraints (e.g. $H: \sigma_1^2 = \sigma_2^2 < \sigma_3^2, \sigma_4^2$, where the comma before σ_4^2 means that no constraint is imposed on this variance). We first apply the FBF to an inequality constrained hypothesis test on the variances of three populations and show that it may not properly incorporate the parsimony introduced by the inequality constraints. We then generalize the aFBF to the problem of testing equality and inequality constrained hypotheses on $J \geqslant 2$ variances. As in Chapter 2, the idea behind the aFBF is that all possible orderings of the variances are equally likely a priori. An application of the aFBF to the inequality constrained hypothesis test shows that it incorporates the parsimony introduced by the inequality constraints. Furthermore, results from a simulation study investigating the performance of the aFBF indicate that it is consistent in the sense that it selects the true hypothesis if the sample size is large enough. We apply the aFBF to empirical data from the Math Garden online learning environment (https://www.mathsgarden.com/) and present a user-friendly software application that can be used to compute the aFBF in an easy manner.

In Chapter 4 we extend the FBF and the BBF to the problem of testing equality and inequality constrained hypotheses on the variances of $J \geqslant 2$ independent populations. As in Chapter 2, the BBF is based on identical priors for the variances, where the hyperparameters of these priors are specified automatically using information from the sample data. In three numerical studies we compared the performance of the FBF, the BBF, and the aFBF as introduced in Chapter 3. We first examined the Bayes factors' behavior when testing nested inequality constrained hypotheses. The results show that the BBF and the aFBF incorporate the parsimony of inequality constrained hypotheses, whereas the FBF may not do so. Next, we investigated information consistency. A Bayes factor is said to be information consistent if it goes to infinity as the effect size goes to infinity, while keeping the sample size fixed. In our numerical study the FBF and the aFBF showed information consistent behavior. The BBF, on the other hand, showed information inconsistent behavior by converging to a constant. Finally, in a simulation study investigating large sample consistency all Bayes factors behaved consistently in the sense that they selected the true hypothesis if the sample size was large enough. Subsequent to the numerical studies we apply the Bayes factors to hypothetical data from four treatment groups as well as to empirical data from two studies investigating attentional performances of Tourette's and ADHD patients and influence of group leaders, respectively.

In Chapter 5 we address the problem of testing inequality constrained hypotheses on the variances of dependent observations (we do not consider equality constraints between the variances in this case for reasons of complexity due to the dependency). In this chapter we apply the encompassing prior approach to computing Bayes factors. In this approach priors under competing inequality constrained hypotheses are formulated as truncations of the prior under the unconstrained hypothesis that does not impose any constraints on the variances. We specify the hyperparameters of this unconstrained prior such that it contains minimal information and all possible orderings of the variances are equally likely a priori. The encompassing prior approach has two main advantages: First, the problem of specifying a prior under every inequality constrained hypothesis to be tested simplifies to specifying one unconstrained prior. Second, computation of the Bayes factor is straightforward using a simple Monte

Carlo method. Our Bayes factor is large sample consistent, which is confirmed in a simulation study investigating the behavior of the Bayes factor when testing an inequality constrained hypothesis against its complement. We apply the Bayes factor to an empirical data set containing repeated measurements of reading recognition in children.

In the epilogue in Chapter 6 we first give a brief summary of the most important aspects of our approach to testing equality and inequality constrained hypotheses on variances and discuss some limitations. Following this, potential directions for future research in the area of testing hypotheses on variances are outlined.

Chapter 2

Automatic Bayes Factors for Testing Variances of Two Independent Normal Distributions

Abstract

Researchers are frequently interested in testing variances of two independent populations. We often would like to know whether the population variances are equal, whether population 1 has smaller variance than population 2, or whether population 1 has larger variance than population 2. In this chapter we consider the Bayes factor, a Bayesian model selection and hypothesis testing criterion, for this multiple hypothesis test. Application of Bayes factors requires specification of prior distributions for the model parameters. Automatic Bayes factors circumvent the difficult task of prior elicitation by using data-driven mechanisms to specify priors in an automatic fashion. In this chapter we develop different automatic Bayes factors for testing two variances: first we apply the fractional Bayes factor (FBF) to the testing problem. It is shown that the FBF does not always function as Occam's razor. Second we develop a new automatic balanced Bayes factor with equal priors for the variances. Third we propose a Bayes factor based on an adjustment of the marginal likelihood in the FBF approach. The latter two methods always function as Occam's razor. Through theoretical considerations and numerical simulations it is shown that the third approach provides strongest evidence in favor of the true hypothesis.

2.1 Introduction

Researchers are frequently interested in comparing two independent populations on a continuous outcome measure. Traditionally, the focus has been on comparing means,

This chapter is published as Böing-Messing, F., & Mulder, J. (2016). Automatic Bayes factors for testing variances of two independent normal distributions. *Journal of Mathematical Psychology*, *72*, 158–170. http://dx.doi.org/10.1016/j.jmp.2015.08.001.

whereas variances are mostly considered nuisance parameters. However, by regarding variances as mere nuisance parameters, one runs the risk of overlooking important information in the data. The variability of a population is a key characteristic which can be the core of a research question. For example, psychological research frequently investigates differences in variability between males and females (e.g. Arden & Plomin, 2006; Borkenau et al., 2013; Feingold, 1992).

In this chapter we consider a Bayesian hypothesis test on the variances of two independent populations. The Bayes factor is a well-known Bayesian criterion for model selection and hypothesis testing (Jeffreys, 1961; Kass & Raftery, 1995). Unlike the p-value, which is often misinterpreted as an error probability (Hubbard & Armstrong, 2006), the Bayes factor has a straightforward interpretation as the relative evidence in the data in favor of a hypothesis as compared to another hypothesis. Moreover, contrary to p-values, the Bayes factor is able to quantify evidence in favor of a null hypothesis (Wagenmakers, 2007). Another useful property, which is not shared by p-values, is that the Bayes factor can straightforwardly be used for testing multiple hypotheses simultaneously (Berger & Mortera, 1999). These and other notions have resulted in a considerable development of Bayes factors for frequently encountered testing problems in the last decade. For example, Klugkist, Laudy, and Hoijtink (2005) proposed Bayes factors for testing analysis of variance models. Rouder, Speckman, Sun, Morey, and Iverson (2009) proposed a Bayesian t-test. Mulder, Hoijtink, and de Leeuw (2012) developed a software program for Bayesian testing of (in)equality constraints on means and regression coefficients in the multivariate normal linear model, and Wetzels and Wagenmakers (2012) proposed Bayesian tests for correlation coefficients. The goal of this chapter is to extend this literature by developing Bayes factors for testing variances. For more interesting references we also refer the reader to the special issue 'Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments' in the *Journal of Mathematical Psychology* in which this chapter appeared (Mulder & Wagenmakers, in preparation).

In applying Bayes factors for hypothesis testing, we need to specify a prior distribution of the model parameters under every hypothesis to be tested. A prior distribution is a probability distribution describing the probability of the possible parameter values before observing the data. In the case of testing two variances, we need to specify a prior for the common variance under the null hypothesis and for the two unique variances under the alternative hypothesis. Specifying priors is a difficult task from a practical point of view, and it is complicated by the fact that we cannot use noninformative improper priors for parameters to be tested because the Bayes factor would then be undefined (Jeffreys, 1961). This has stimulated researchers to develop Bayes factors which do not require prior elicitation using external prior information. Instead, these so-called automatic Bayes factors use information from the sample data to specify priors in an automatic fashion. So far, however, no automatic Bayes factors have been developed for testing variances.

In this chapter we develop three types of automatic Bayes factors for testing variances of two independent normal populations. We first consider the fractional Bayes factor (FBF) of O'Hagan (1995) and apply it for the first time to the problem of testing variances. In the FBF methodology the likelihood of the complete data is divided into two fractions: one for specifying the prior and one for testing the hypotheses. However, it has been shown (e.g. Mulder, 2014b) that the FBF may not

be suitable for testing inequality constrained hypotheses (e.g. variance 1 is smaller than variance 2) because it may not function as Occam's razor. In other words, the FBF may not prefer the simpler hypothesis when two hypotheses fit the data equally well. This is a consequence of the fact that in the FBF the automatic prior is located at the likelihood of the data. We develop two novel solutions to this problem: the first is an automatic Bayes factor with equal automatic priors for both variances under the alternative hypothesis. This methodology is related to the constrained posterior priors approach of Mulder, Hoijtink, and Klugkist (2010). The second novel solution is an automatic Bayes factor based on adjusting the definition of the FBF such that the resulting automatic Bayes factor always functions as Occam's razor. This approach is related to the work of Mulder (2014b), with the difference that our method results in stronger evidence in favor of a true null hypothesis.

The remainder of this chapter is structured as follows. In the next section we provide details on the normal model to be used and introduce the hypotheses we shall be concerned with. We then discuss five theoretical properties which are used for evaluating the automatic Bayes factors. Following this, we develop the three automatic Bayes factors and evaluate them according to the theoretical properties. Subsequently, the performance of the Bayes factors is investigated by means of a small simulation study. We conclude the chapter with an application of the Bayes factors to two empirical data examples and a discussion of possible extensions and limitations of our approaches.

2.2 Model and Hypotheses

We assume that the outcome variable of interest, X, is normally distributed in both populations:

$$X_j \sim \mathrm{N}\left(\mu_j, \sigma_j^2\right), \quad j = 1, 2, \tag{2.1}$$

where j is the population index and μ_j and σ_j^2 are the population-specific parameters. The unknown parameter in this model is $\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)' = \left((\mu_1, \mu_2,)', \left(\sigma_1^2, \sigma_2^2\right)'\right)' \in \mathbb{R}^2 \times \Omega_u$, where $\Omega_u := \left(\mathbb{R}^+\right)^2$ is the unconstrained parameter space of $\boldsymbol{\sigma}^2$.

In this chapter we shall be concerned with testing the following nonnested (in)equality constrained hypotheses against one another:

$$\begin{array}{lcl} H_0\colon \sigma_1^2 = \sigma_2^2 = \sigma^2, & & H_0\colon \sigma^2 \in \Omega_0 := \mathbb{R}^+, \\ H_1\colon \sigma_1^2 < \sigma_2^2, & \Leftrightarrow & H_1\colon \boldsymbol{\sigma}^2 \in \Omega_1 := \left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 < \sigma_2^2\right\}, \\ H_2\colon \sigma_1^2 > \sigma_2^2, & & H_2\colon \boldsymbol{\sigma}^2 \in \Omega_2 := \left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 > \sigma_2^2\right\}, \end{array} \tag{2.2}$$

where $\Omega_1, \Omega_2 \subset \Omega_u$ and Ω_0 denote the parameter spaces under the corresponding (in)equality constrained hypotheses.

We made two choices in formulating the hypotheses in Equation (2.2). First, we do not test any constraints on the mean parameters μ_1 and μ_2. This is because the objective of this chapter is to provide a Bayesian alternative to the classical frequentist procedures for testing two variances. For a general framework for testing (in)equality constrained hypotheses on mean parameters, see, for example, Mulder et al. (2012). The second choice we made is to divide the classical alternative hypothesis

$H_a\colon \sigma_1^2 \neq \sigma_2^2 \Leftrightarrow H_a\colon \sigma_1^2 < \sigma_2^2 \vee \sigma_1^2 > \sigma_2^2$ into two separate hypotheses, $H_1\colon \sigma_1^2 < \sigma_2^2$ and $H_2\colon \sigma_1^2 > \sigma_2^2$ ($\vee$ denotes logical disjunction and reads "or"). The advantage of this approach is that it allows us to quantify and compare the evidence in favor of a negative effect (H_1) and a positive effect (H_2). This is of great interest to applied researchers, who would often like to know not only whether there is an effect, but also in what direction.

Another hypothesis we will consider is the unconstrained hypothesis

$$H_u\colon \sigma_1^2, \sigma_2^2 > 0 \quad \Leftrightarrow \quad H_u\colon \boldsymbol{\sigma}^2 \in \Omega_u = \left(\mathbb{R}^+\right)^2. \tag{2.3}$$

This hypothesis is not of substantial interest to us because it is entirely covered by the hypotheses in Equation (2.2). In other words, $\{H_0, H_1, H_2\}$ is a partition of H_u. The unconstrained hypothesis will be used to evaluate theoretical properties of the priors and Bayes factors such as balancedness and Occam's razor (discussed in the next section).

2.3 Properties for the Automatic Priors and Bayes Factors

Based on the existing literature on automatic Bayes factors, we shall focus on the following theoretical properties when evaluating the automatic priors and Bayes factors:

1. *Proper priors: The priors must be proper probability distributions.* When using improper priors on parameters that are tested, the resulting Bayes factors depend on unspecified constants (see, for instance, O'Hagan, 1995). Improper priors may only be used on common nuisance parameters that are present under all hypotheses to be tested (Jeffreys, 1961).

2. *Minimal information: Priors under composite hypotheses should contain the information of a minimal study.* Using arbitrarily vague priors gives rise to the Jeffreys–Lindley paradox (Jeffreys, 1961; Lindley, 1957), whereas priors containing too much information about the parameters will dominate the data. Therefore it is often suggested to let the prior contain the information of a minimal study (e.g. Berger & Pericchi, 1996; O'Hagan, 1995; Spiegelhalter & Smith, 1982). A minimal study is the smallest possible study (in terms of sample size) for which all free parameters under all hypotheses are identifiable. If prior information is absent (as is usually the case when automatic Bayes factors are considered), then a prior containing minimal information is a reasonable starting point.

3. *Scale invariance: The Bayes factors should be invariant under rescaling of the data.* In other words, the Bayes factors should not depend on the scale of the outcome variable. This is important because when comparing, say, the heterogeneity of ability scores of males and females, it should not matter if the ability test has a scale from 0 to 10 or from 0 to 100.

4. *Balancedness: The prior under the unconstrained hypothesis should be balanced.* If we denote $\eta = \log\left(\sigma_1^2/\sigma_2^2\right)$, then the unconstrained hypothesis can be written

as $H_u\colon \eta \in \mathbb{R}$. The prior for η under H_u should be symmetric about 0 and nonincreasing in $|\eta|$ (e.g. Berger & Delampady, 1987). Following Jeffreys (1961), we shall refer to a prior satisfying these properties as a balanced prior. A balanced prior can be considered objective in two respects: first, the symmetry ensures that neither a positive nor a negative effect is preferred a priori. Second, the nonincreasingness ensures that no other values but 0 are treated as special.

5. *Occam's razor: The Bayes factors should function as Occam's razor.* Occam's razor is the principle that if two hypotheses fit the data equally well, then the simpler (i.e. less complex) hypothesis should be preferred. The principle is based on the empirical observation that simple hypotheses that fit the data are more likely to be correct than complicated ones. When testing nested hypotheses, Bayes factors automatically function as Occam's razor by balancing fit and complexity of the hypotheses (Kass & Raftery, 1995). When testing inequality constrained hypotheses, however, the Bayes factor does not always function as Occam's razor (Mulder, 2014a).

2.4 Automatic Bayes Factors

The Bayes factor is a Bayesian hypothesis testing criterion that is related to the likelihood ratio statistic. It is equal to the ratio of the marginal likelihoods under two competing hypotheses:

$$B_{pq} = \frac{m_p\left(\mathbf{x}\right)}{m_q\left(\mathbf{x}\right)}, \tag{2.4}$$

where B_{pq} denotes the Bayes factor comparing hypotheses H_p and H_q, and $m_p\left(\mathbf{x}\right)$ is the marginal likelihood under hypothesis H_p as a function of the data $\mathbf{x}$.

2.4.1 Fractional Bayes Factor

The fractional Bayes factor introduced by O'Hagan (1995) is a general, automatic method for comparing two statistical models or hypotheses. In this chapter we apply it for the first time to the problem of testing variances. We use the superscript F to refer to the FBF.

Marginal Likelihoods

The FBF marginal likelihood under hypothesis H_p, $p = 0, 1, 2, u$, is given by

$$m_p^F\left(\mathbf{b}, \mathbf{x}\right) = \frac{\int_{\Omega_p}\int_{\mathbb{R}^2} f_p\left(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)\pi_p^N\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}{\int_{\Omega_p}\int_{\mathbb{R}^2} f_p\left(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)^{\mathbf{b}}\pi_p^N\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}, \tag{2.5}$$

where $p = u$ refers to the unconstrained hypothesis (with a slight abuse of notation), and under H_0 the variance parameter $\boldsymbol{\sigma}^2$ is a scalar containing only the common variance σ^2. Here $\pi_p^N\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)$ is the noninformative Jeffreys prior on $\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)'$. Under H_0 it is $\pi_0^N\left(\boldsymbol{\mu},\sigma^2\right) \propto \sigma^{-2}$, while under H_u we have $\pi_u^N\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2\right) \propto \sigma_1^{-2}\sigma_2^{-2}$. Under H_p, $p = 1, 2$, the Jeffreys prior is $\pi_p^N\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2\right) \propto \sigma_1^{-2}\sigma_2^{-2}1_{\Omega_p}\left(\boldsymbol{\sigma}^2\right)$, where $1_{\Omega_p}\left(\boldsymbol{\sigma}^2\right)$ is the

indicator function which is 1 if $\boldsymbol{\sigma}^2 \in \Omega_p$ and 0 otherwise. The expression $f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}}$ denotes a fraction of the likelihood, the cornerstone of the FBF methodology. Let $\mathbf{x}_j = \left(x_{1j}, \ldots, x_{n_j j}\right)'$ be a vector of n_j observations coming from X_j. Fractions of the likelihoods under the four hypotheses are given by

$$\begin{aligned} f_0\left(\mathbf{x}|\boldsymbol{\mu}, \sigma^2\right)^{\mathbf{b}} &:= f\left(\mathbf{x}_1|\mu_1, \sigma^2\right)^{b_1} f\left(\mathbf{x}_2|\mu_2, \sigma^2\right)^{b_2}, \\ f_u\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}} &:= f\left(\mathbf{x}_1|\mu_1, \sigma_1^2\right)^{b_1} f\left(\mathbf{x}_2|\mu_2, \sigma_2^2\right)^{b_2}, \\ f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}} &:= f_u\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}} 1_{\Omega_p}\left(\boldsymbol{\sigma}^2\right), \quad p = 1, 2, \end{aligned} \tag{2.6}$$

where

$$f\left(\mathbf{x}_j|\mu_j, \sigma_j^2\right)^{b_j} = \left(\prod_{i=1}^{n_j} \mathrm{N}\left(x_{ij}|\mu_j, \sigma_j^2\right)\right)^{b_j} \tag{2.7}$$

is a fraction of the likelihood of population j (e.g. Berger & Pericchi, 2001). Here $b_1 \in (1/n_1, 1]$ and $b_2 \in (1/n_2, 1]$ are population-specific proportions to be determined by the user, and by using $\mathbf{b} = (b_1, b_2)'$ as a superscript we slightly abuse notation. We obtain the full likelihood $f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)$ by setting $b_1 = b_2 = 1$.

Plugging $f_0\left(\mathbf{x}|\boldsymbol{\mu}, \sigma^2\right)$, $f_0\left(\mathbf{x}|\boldsymbol{\mu}, \sigma^2\right)^{\mathbf{b}}$, and $\pi_0^N\left(\boldsymbol{\mu}, \sigma^2\right)$ into Equation (2.5), we obtain the marginal likelihood under H_0 after some algebra (see Appendix 2.A) as

$$m_0^F(\mathbf{b}, \mathbf{x}) = \frac{(b_1 b_2)^{\frac{1}{2}} \Gamma\left(\frac{n_1+n_2-2}{2}\right)\left(b_1(n_1-1)s_1^2 + b_2(n_2-1)s_2^2\right)^{\frac{b_1 n_1 + b_2 n_2 - 2}{2}}}{\pi^{\frac{n_1(1-b_1)+n_2(1-b_2)}{2}} \Gamma\left(\frac{b_1 n_1 + b_2 n_2 - 2}{2}\right)\left((n_1-1)s_1^2 + (n_2-1)s_2^2\right)^{\frac{n_1+n_2-2}{2}}}, \tag{2.8}$$

where Γ denotes the gamma function, and $s_j^2 = \frac{1}{n_j - 1}\sum_{i=1}^{n_j}\left(x_{ij} - \bar{x}_j\right)^2$ is the sample variance of $\mathbf{x}_j$, $j = 1, 2$. The marginal likelihoods under H_1 and H_2 are functions of the marginal likelihood under H_u, which is given by

$$m_u^F(\mathbf{b}, \mathbf{x}) = \frac{\pi^{-\frac{n_1(1-b_1)+n_2(1-b_2)}{2}} b_1^{\frac{b_1 n_1}{2}} b_2^{\frac{b_2 n_2}{2}} \Gamma\left(\frac{n_1-1}{2}\right)\Gamma\left(\frac{n_2-1}{2}\right)}{\Gamma\left(\frac{b_1 n_1 - 1}{2}\right)\Gamma\left(\frac{b_2 n_2 - 1}{2}\right)\left((n_1-1)s_1^2\right)^{\frac{n_1(1-b_1)}{2}}\left((n_2-1)s_2^2\right)^{\frac{n_2(1-b_2)}{2}}}. \tag{2.9}$$

For the marginal likelihoods under H_1 and H_2 we then have

$$m_p^F(\mathbf{b}, \mathbf{x}) = \frac{P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}\right)}{P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}^{\mathbf{b}}\right)} m_u^F(\mathbf{b}, \mathbf{x}), \quad p = 1, 2. \tag{2.10}$$

Here $P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}\right)$ and $P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}^{\mathbf{b}}\right)$ denote the probability that $\boldsymbol{\sigma}^2$ is in Ω_p given the complete data $\mathbf{x}$ or a fraction thereof (for which we use the notation $\mathbf{x}^{\mathbf{b}}$). The exact expressions for the two probabilities are given in Equations (2.33) and (2.34) in Appendix 2.B. The derivation of Equations (2.9) and (2.10) is analogous to that of Equation (2.8) given in Appendix 2.A.

Evaluation of the Method

We will now evaluate the FBF according to the five properties discussed in Section 2.3:

1. *Proper priors.* First, note that the marginal likelihood in Equation (2.5) can be rewritten as

$$\begin{aligned} & m_p^F(\mathbf{b}, \mathbf{x}) \\ &= \int_{\Omega_p} \int_{\mathbb{R}^2} f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{1}-\mathbf{b}} \frac{f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}} \pi_p^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)}{\int_{\Omega_p} \int_{\mathbb{R}^2} f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}} \pi_p^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) d\boldsymbol{\mu} d\boldsymbol{\sigma}^2} d\boldsymbol{\mu} d\boldsymbol{\sigma}^2 \\ &= \int_{\Omega_p} \int_{\mathbb{R}^2} f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{1}-\mathbf{b}} \pi_p^F\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) d\boldsymbol{\mu} d\boldsymbol{\sigma}^2, \end{aligned} \tag{2.11}$$

   where we use the superscript $\mathbf{1} - \mathbf{b} = (1 - b_1, 1 - b_2)'$ analogously to $\mathbf{b}$ in Equation (2.6). Here $\pi_p^F\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) \propto f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}} \pi_p^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)$ is a posterior prior obtained by updating the Jeffreys prior with a fraction of the likelihood. It can be considered the automatic prior implied by the FBF approach and is proper if $b_1 n_1 + b_2 n_2 > 2$ under H_0 and $b_j n_j > 1$, $j = 1, 2$, under H_1, H_2, and H_u. We use the notation $\mathbf{x}^{\mathbf{b}}$ to indicate that it is based on a fraction $\mathbf{b}$ of the likelihood of the complete sample data $\mathbf{x}$.

2. *Minimal information.* A minimal study consists of four observations, two from each population. This is because we need two observations from population j for $\left(\mu_j, \sigma_j^2\right)'$ to be identifiable. We can make the priors contain the information of a minimal study by setting $\mathbf{b} = (2/n_1, 2/n_2)'$ (O'Hagan, 1995).

3. *Scale invariance.* Multiplying all observations in $\mathbf{x}_j$ by a constant w results in a sample variance of $w^2 s_j^2$, $j = 1, 2$. Plugging $w^2 s_j^2$ into the formulas for the marginal likelihoods in Equations (2.8) and (2.9) does not change the resulting Bayes factors. Thus the FBF is scale invariant.

4. *Balancedness.* The marginal unconstrained prior on $\boldsymbol{\sigma}^2$ implied by the FBF approach is given by

$$\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) = \text{Inv-}\chi^2\left(\sigma_1^2|\nu_1, \tau_1^2\right) \text{Inv-}\chi^2\left(\sigma_2^2|\nu_2, \tau_2^2\right), \tag{2.12}$$

   where

$$\nu_j = b_j n_j - 1 \quad \text{and} \quad \tau_j^2 = \frac{b_j\left(n_j - 1\right) s_j^2}{b_j n_j - 1}, \quad j = 1, 2. \tag{2.13}$$

   Here Inv-$\chi^2\left(\nu, \tau^2\right)$ is the scaled inverse-χ^2 distribution with degrees of freedom hyperparameter $\nu > 0$ and scale hyperparameter $\tau^2 > 0$ (Gelman, Carlin, Stern, & Rubin, 2004). The corresponding unconstrained prior on $\eta = \log\left(\sigma_1^2/\sigma_2^2\right)$, $\pi_u^F(\eta|\mathbf{x}^{\mathbf{b}})$, is balanced if and only if $\nu_1 = \nu_2 \wedge \tau_1^2 = \tau_2^2$ ($\wedge$ denotes logical conjunction and reads "and"; see Appendix 2.C for a proof). In practice the sample sizes and sample variances will commonly be such that $\neg\left(\nu_1 = \nu_2 \wedge \tau_1^2 = \tau_2^2\right)$, which is why $\pi_u^F(\eta|\mathbf{x}^{\mathbf{b}})$ will commonly be unbalanced ($\neg$ denotes logical negation and reads "not"). Figure 2.1 illustrates this. The figure shows the priors on $\boldsymbol{\sigma}^2$ (top row) and η (bottom row) for sample variances $s_1^2 = 1$ and $s_2^2 \in \{1, 4, 16\}$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.1$. It can be seen that $\pi_u^F(\eta|\mathbf{x}^{\mathbf{b}})$ is only balanced if $s_2^2 = s_1^2 = 1$, in which case $\nu_1 = \nu_2 \wedge \tau_1^2 = \tau_2^2$. For $s_2^2 \in \{4, 16\}$ it is shifted to the left (i.e. it is not skewed).

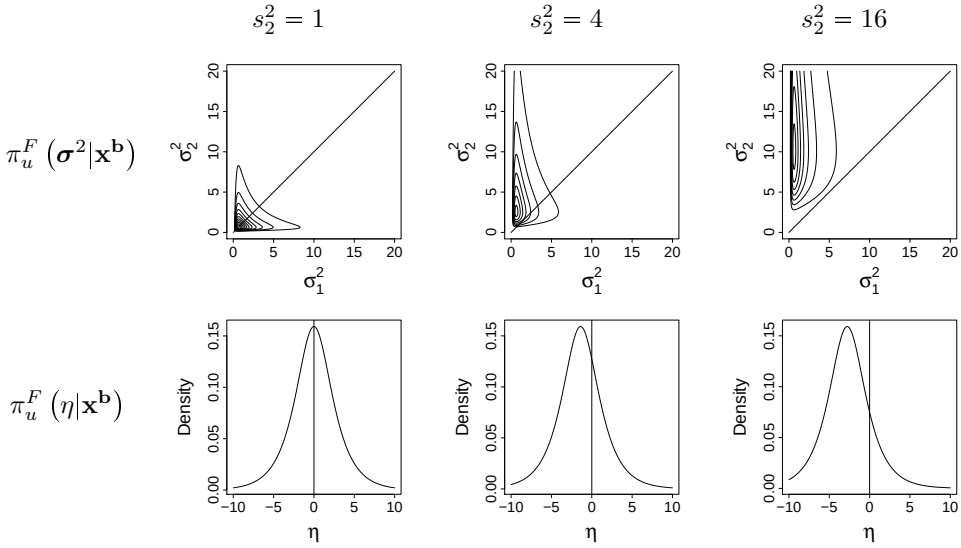


Figure 2.1: The marginal unconstrained FBF prior $\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x^b}\right)$ (top row) and the corresponding prior $\pi_u^F\left(\eta = \log\left(\sigma_1^2/\sigma_2^2\right)|\mathbf{x^b}\right)$ (bottom row) for sample variances $s_1^2 = 1$ and $s_2^2 \in \{1, 4, 16\}$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.1$. The prior $\pi_u^F(\eta|\mathbf{x^b})$ is only balanced when $s_2^2 = s_1^2 = 1$.

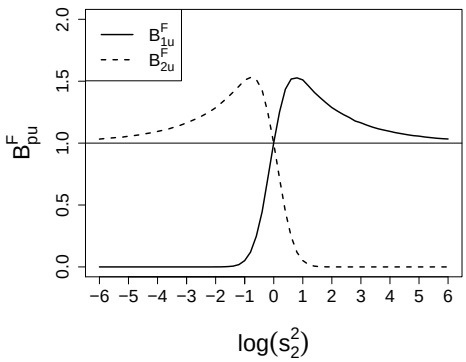


Figure 2.2: Bayes factors B_{1u}^F (solid line) and B_{2u}^F (dashed line) for sample variances $s_1^2 = 1$ and $s_2^2 \in [\exp(-6), \exp(6)]$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.1$. The Bayes factors approach 1 for very large and very small s_2^2, respectively. That is, they do not favor the more parsimonious inequality constrained hypothesis even though it is strongly supported by the data. This shows that B_{1u}^F and B_{2u}^F do not function as Occam's razor.

5. *Occam's razor.* The Bayes factors B^F_{1u} and B^F_{2u} should function as Occam's razor by favoring the simplest hypothesis that is in line with the data. This, however, is not the case, as Figure 2.2 illustrates. The plot shows B^F_{1u} (solid line) and B^F_{2u} (dashed line) for sample variances $s_1^2 = 1$ and $s_2^2 \in [\exp(-6), \exp(6)]$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.1$. It can be seen that B^F_{1u} and B^F_{2u} approach 1 for very large and very small s_2^2, respectively. Thus B^F_{1u} and B^F_{2u} are indecisive despite the data strongly supporting the more parsimonious inequality constrained hypothesis. This undesirable property is a direct consequence of the fact that the unconstrained prior is located at the likelihood of the data.

2.4.2 Balanced Bayes Factor

In the previous section we have seen that the FBF involves two problems: the marginal unconstrained prior $\pi^F_u\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)$ is unbalanced and the Bayes factors B^F_{pu} and B^F_{p0}, $p = 1, 2$, do not function as Occam's razor. In this section we propose a solution to these problems which we refer to as the balanced Bayes factor (BBF). The BBF is a new automatic Bayes factor for testing variances of two independent normal distributions that satisfies all five properties discussed in Section 2.3. The BBF approach is related to the constrained posterior priors approach of Mulder et al. (2010) with the exception that the latter uses empirical training samples for prior specification instead of a fraction of the likelihood. The fractional approach of the BBF is therefore computationally less demanding. We use the superscript B to refer to the BBF.

Marginal Likelihoods

In the FBF approach the marginal unconstrained prior $\pi^F_u\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) = \text{Inv-}\chi^2\left(\sigma_1^2|\nu_1, \tau_1^2\right)\text{Inv-}\chi^2\left(\sigma_2^2|\nu_2, \tau_2^2\right)$ is balanced if and only if $\nu_1 = \nu_2 \wedge \tau_1^2 = \tau_2^2$, which in practice will rarely be the case. The main idea of the BBF thus is to replace $\pi^F_u\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)$ with a marginal unconstrained prior $\pi^B_u\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) = \text{Inv-}\chi^2\left(\sigma_1^2|\nu, \tau^2\right) \times \text{Inv-}\chi^2\left(\sigma_2^2|\nu, \tau^2\right)$ with common hyperparameters ν and τ^2. This way $\pi^B_u\left(\eta|\mathbf{x}^{\mathbf{b}}\right)$ is balanced by definition (see Appendix 2.C). As with the FBF, we shall use information from the sample data $\mathbf{x}$ to define ν and τ^2: first we assume that $\sigma_1^2 = \sigma_2^2$ and update the Jeffreys prior with a fraction of the likelihood under H_0, $f_0\left(\mathbf{x}|\boldsymbol{\mu}, \sigma^2\right)^{\mathbf{b}}$. Note that this results in the FBF posterior prior $\pi^F_0\left(\boldsymbol{\mu}, \sigma^2|\mathbf{x}^{\mathbf{b}}\right)$. Next, we obtain the marginal posterior prior on σ^2 by integrating out $\boldsymbol{\mu}$:

$$\pi^F_0\left(\sigma^2|\mathbf{x}^{\mathbf{b}}\right) = \int_{\mathbb{R}^2} \pi^F_0\left(\boldsymbol{\mu}, \sigma^2|\mathbf{x}^{\mathbf{b}}\right) d\boldsymbol{\mu} = \text{Inv-}\chi^2\left(\sigma^2|\nu_\bullet, \tau_\bullet^2\right), \tag{2.14}$$

where

$$\nu_\bullet = b_1 n_1 + b_2 n_2 - 2 \quad \text{and} \quad \tau_\bullet^2 = \frac{b_1\left(n_1 - 1\right)s_1^2 + b_2\left(n_2 - 1\right)s_2^2}{b_1 n_1 + b_2 n_2 - 2}. \tag{2.15}$$

We use the subscript $\bullet$ to indicate that the hyperparameters $\nu_\bullet$ and $\tau_\bullet^2$ combine information from both samples $\mathbf{x}_1$ and $\mathbf{x}_2$. We propose using the distribution in Equation (2.14) as the prior on both σ_1^2 and σ_2^2 under H_u, giving us the BBF marginal

unconstrained prior on $\boldsymbol{\sigma}^2$ as

$$\pi_u^B\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) = \pi_0^F\left(\sigma_1^2|\mathbf{x}^{\mathbf{b}}\right)\pi_0^F\left(\sigma_2^2|\mathbf{x}^{\mathbf{b}}\right), \tag{2.16}$$

with $\pi_0^F\left(\sigma_j^2|\mathbf{x}^{\mathbf{b}}\right)$ as in Equation (2.14). Note that b_1 and b_2 need to be specified such that $b_1 n_1 + b_2 n_2 > 2$ for $\nu_\bullet$ to be positive. With the marginal unconstrained prior at hand, we define the joint prior on $\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)'$ under H_u as

$$\pi_u^B\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) = \pi_u^B\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)\pi^N\left(\boldsymbol{\mu}\right), \tag{2.17}$$

with $\pi_u^B\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)$ as in Equation (2.16). Here $\pi^N\left(\boldsymbol{\mu}\right) \propto 1$ is the Jeffreys prior for $\boldsymbol{\mu}$, which we may use since in our testing problem $\boldsymbol{\mu}$ is a common nuisance parameter that is present under all hypotheses. We shall define the BBF priors under H_1 and H_2 as truncations of the prior under H_u (Berger & Mortera, 1999; Klugkist, Laudy, & Hoijtink, 2005):

$$\begin{aligned}\pi_p^B\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) &= \frac{1}{P^B\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}^{\mathbf{b}}\right)}\pi_u^B\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)1_{\Omega_p}\left(\boldsymbol{\sigma}^2\right) \\ &= 2 \cdot \pi_u^B\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)1_{\Omega_p}\left(\boldsymbol{\sigma}^2\right), \quad p = 1, 2,\end{aligned} \tag{2.18}$$

where

$$P^B\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}^{\mathbf{b}}\right) = \int_{\Omega_p}\int_{\mathbb{R}^2}\pi_u^B\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2 = \int_{\Omega_p}\pi_u^B\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)d\boldsymbol{\sigma}^2 = 0.5. \tag{2.19}$$

We have $P^B\left(\boldsymbol{\sigma}^2 \in \Omega_1|\mathbf{x}^{\mathbf{b}}\right) = P^B\left(\boldsymbol{\sigma}^2 \in \Omega_2|\mathbf{x}^{\mathbf{b}}\right) = 0.5$ because $\pi_u^B\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)$ is the product of two identical scaled inverse-χ^2 distributions. In Equation (2.18) the inverse $1/P^B\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}^{\mathbf{b}}\right)$ acts as a normalizing constant. Eventually, we define the BBF prior under H_0 such that it is in line with the priors under H_1 and H_2:

$$\pi_0^B\left(\boldsymbol{\mu}, \sigma^2|\mathbf{x}^{\mathbf{b}}\right) = \pi_0^F\left(\sigma^2|\mathbf{x}^{\mathbf{b}}\right)\pi^N\left(\boldsymbol{\mu}\right), \tag{2.20}$$

with $\pi_0^F\left(\sigma^2|\mathbf{x}^{\mathbf{b}}\right)$ as in Equation (2.14).

With the priors at hand we can now determine the marginal likelihoods. The BBF marginal likelihood under hypothesis H_p, $p = 0, 1, 2, u$, is given by

$$m_p^B\left(\mathbf{b}, \mathbf{x}\right) = \int_{\Omega_p}\int_{\mathbb{R}^2} f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)\pi_p^B\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2. \tag{2.21}$$

Besides the prior, this formulation differs from the FBF marginal likelihood in another important aspect: in Equation (2.11) we have seen that to compute the FBF marginal likelihood we implicitly factor the full likelihood as $f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) = f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{1-b}} \times f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}}$. Then a proper posterior prior is obtained using $f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}}$, and the marginal likelihood is computed using the remaining fraction $f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{1-b}}$. From Equation (2.21) it can be seen that to compute the BBF marginal likelihoods we use the full likelihood $f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)$ instead of $f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{1-b}}$. That is, we first use $f_0\left(\mathbf{x}|\boldsymbol{\mu}, \sigma^2\right)^{\mathbf{b}}$ to obtain the proper prior $\pi_u^B\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)$, and subsequently we use $f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)$ to compute the marginal likelihoods. This implies that we use the data

twice, once for prior specification and once for hypothesis testing. We choose to do so for the following reason: we use the information in $f_0\left(\mathbf{x}|\boldsymbol{\mu},\sigma^2\right)^{\mathbf{b}}$ to specify the variance of the balanced prior, but not its location. This means that we use less information for prior specification than is actually contained in $f_0\left(\mathbf{x}|\boldsymbol{\mu},\sigma^2\right)^{\mathbf{b}}$. Therefore, the full likelihood $f_p\left(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)$ is used for hypothesis testing. The latter illustrates that the BBF approach differs fundamentally from standard automatic procedures such as the FBF in which the likelihood is explicitly divided into a training part and a testing part. This is reflected in the function of $\mathbf{b}$ in the FBF and the BBF: while in the FBF the $\mathbf{b}$ determines how the likelihood is divided, in the BBF it determines how much of the information in the data we want to use twice.

Now, plugging $f_0\left(\mathbf{x}|\boldsymbol{\mu},\sigma^2\right)$ and $\pi_0^B\left(\boldsymbol{\mu},\sigma^2|\mathbf{x}^{\mathbf{b}}\right)$ into Equation (2.21), we obtain the BBF marginal likelihood under H_0 as

$$m_0^B\left(\mathbf{b},\mathbf{x}\right)=\frac{k\left(\nu_\bullet\tau_\bullet^2\right)^{\frac{\nu_\bullet}{2}}\Gamma\left(\frac{n_1+n_2+\nu_\bullet-2}{2}\right)}{\pi^{\frac{n_1+n_2-2}{2}}\Gamma\left(\frac{\nu_\bullet}{2}\right)\left(n_1n_2\right)^{\frac{1}{2}}\left(\left(n_1-1\right)s_1^2+\left(n_2-1\right)s_2^2+\nu_\bullet\tau_\bullet^2\right)^{\frac{n_1+n_2+\nu_\bullet-2}{2}}}, \tag{2.22}$$

with $\nu_\bullet$ and $\tau_\bullet^2$ as in Equation (2.15), and k is an unspecified constant coming from the improper Jeffreys prior on the common mean parameter, $\pi^N\left(\boldsymbol{\mu}\right)$ (similar to k_0 in Appendix 2.A).

The marginal likelihoods under H_1 and H_2 are functions of the marginal likelihood under H_u, which is

$$m_u^B\left(\mathbf{b},\mathbf{x}\right)=\frac{k\,\pi^{-\frac{n_1+n_2-2}{2}}\left(n_1n_2\right)^{-\frac{1}{2}}\left(\nu_\bullet\tau_\bullet^2\right)^{\nu_\bullet}\Gamma\left(\frac{n_1+\nu_\bullet-1}{2}\right)\Gamma\left(\frac{n_2+\nu_\bullet-1}{2}\right)}{\Gamma\left(\frac{\nu_\bullet}{2}\right)^2\left(\left(n_1-1\right)s_1^2+\nu_\bullet\tau_\bullet^2\right)^{\frac{n_1+\nu_\bullet-1}{2}}\left(\left(n_2-1\right)s_2^2+\nu_\bullet\tau_\bullet^2\right)^{\frac{n_2+\nu_\bullet-1}{2}}}, \tag{2.23}$$

with k as in Equation (2.22). The marginal likelihoods under H_1 and H_2 are then given by

$$m_p^B\left(\mathbf{b},\mathbf{x}\right)=\frac{P^B\left(\boldsymbol{\sigma}^2\in\Omega_p|\mathbf{x}\right)}{P^B\left(\boldsymbol{\sigma}^2\in\Omega_p|\mathbf{x}^{\mathbf{b}}\right)}m_u^B\left(\mathbf{b},\mathbf{x}\right)=2\cdot P^B\left(\boldsymbol{\sigma}^2\in\Omega_p|\mathbf{x}\right)\cdot m_u^B\left(\mathbf{b},\mathbf{x}\right),\quad p=1,2, \tag{2.24}$$

with $P^B\left(\boldsymbol{\sigma}^2\in\Omega_p|\mathbf{x}^{\mathbf{b}}\right)$ as in Equation (2.19), and the exact expression for $P^B\left(\boldsymbol{\sigma}^2\in\Omega_p|\mathbf{x}\right)$ is given in Equation (2.35) in Appendix 2.B. The derivation of Equations (2.22), (2.23) and (2.24) follows steps similar to those in Appendix 2.A. Note that the unspecified constant k cancels out in the computation of Bayes factors.

Evaluation of the Method

We will now evaluate the BBF according to the five properties discussed in Section 2.3:

1. *Proper priors.* Equations (2.18) and (2.20), in combination with Equations (2.14)–(2.17), show that the priors on $\boldsymbol{\sigma}^2$ under H_0, H_1, and H_2 are proper (truncated) scaled-inverse-χ^2 distributions if $b_1n_1+b_2n_2>2$.

2. *Minimal information.* As was set out in the previous section, the unconstrained prior is based on the assumption that $\sigma_1^2=\sigma_2^2=\sigma^2$. A minimal study therefore

consists of three observations, with at least one observation from each population. We can thus make the priors contain the information of a minimal study by setting $\mathbf{b} = (1.5/n_1, 1.5/n_2)'$. Note that this results in degrees of freedom of $\nu_\bullet = 1$ (see Equation (2.15)).

3. *Scale invariance.* The BBF is scale-invariant for the same reason that the FBF is (see Section 2.4.1).

4. *Balancedness.* As was mentioned before, the unconstrained prior $\pi_u^B\left(\eta|\mathbf{x}^{\mathbf{b}}\right)$ is balanced by definition. An illustration is given in Figure 2.3, which shows the priors on $\boldsymbol{\sigma}^2$ (top row) and η (bottom row) for sample variances $s_1^2 = 1$ and $s_2^2 \in \{1, 4, 16\}$, sample sizes $n_1 = n_2 = 20 = n$, and fractions $b_1 = b_2 = 1.5/n = 1.5/20 = 0.075$. It can be seen that $\pi_u^B\left(\eta|\mathbf{x}^{\mathbf{b}}\right)$ is always balanced.

5. *Occam's razor.* Figure 2.4 shows the Bayes factors B_{1u}^B (solid line) and B_{2u}^B (dashed line) for sample variances $s_1^2 = 1$ and $s_2^2 \in [\exp(-6), \exp(6)]$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.075$. It can be seen that B_{1u}^B (B_{2u}^B) increases (decreases) monotonically as s_2^2 increases, favoring the more parsimonious inequality constrained hypothesis over the unconstrained hypothesis if the former is supported by the data. The Bayes factors thus function as Occam's razor. In fact, the Bayes factors go to 2 for very large and very small s_2^2, respectively, because H_1 and H_2 are twice as parsimonious as H_u.

2.4.3 Adjusted Fractional Bayes Factor

Mulder (2014b) proposed a modification of the integration region in the FBF marginal likelihood under (in)equality constrained hypotheses to ensure that the latter always incorporates the complexity of an inequality constrained hypothesis. Compared to the FBF, the proposed modification is always larger for an inequality constrained hypothesis that is supported by the data. Even though this is essentially a good property, a possible disadvantage of this approach is that it results in a slight decrease of the evidence in favor of a true null hypothesis. For this reason we propose an alternative method in this chapter: we adjust the FBF marginal likelihood under an inequality constrained hypothesis as suggested by Mulder (2014b), but we keep the marginal likelihood under the equality constrained hypothesis as in the FBF approach. We shall refer to this approach as the adjusted fractional Bayes factor (aFBF). We use the superscript aF to refer to the aFBF.

Marginal Likelihoods

Following Mulder (2014b), we define the adjusted FBF marginal likelihood under an inequality constrained hypothesis as

$$m_p^{aF}(\mathbf{b}, \mathbf{x}) = \frac{\int_{\Omega_p}\int_{\mathbb{R}^2} f_u\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)\pi_u^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}{\int_{\Omega_p^a}\int_{\mathbb{R}^2} f_u\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}}\pi_u^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}, \quad p = 1, 2, \tag{2.25}$$

where $\mathbf{b} = (b_1, b_2)' \in (1/n_1, 1] \times (1/n_2, 1]$ as with the FBF. Note the two adjustments that were made compared to the standard FBF marginal likelihood given in Equation

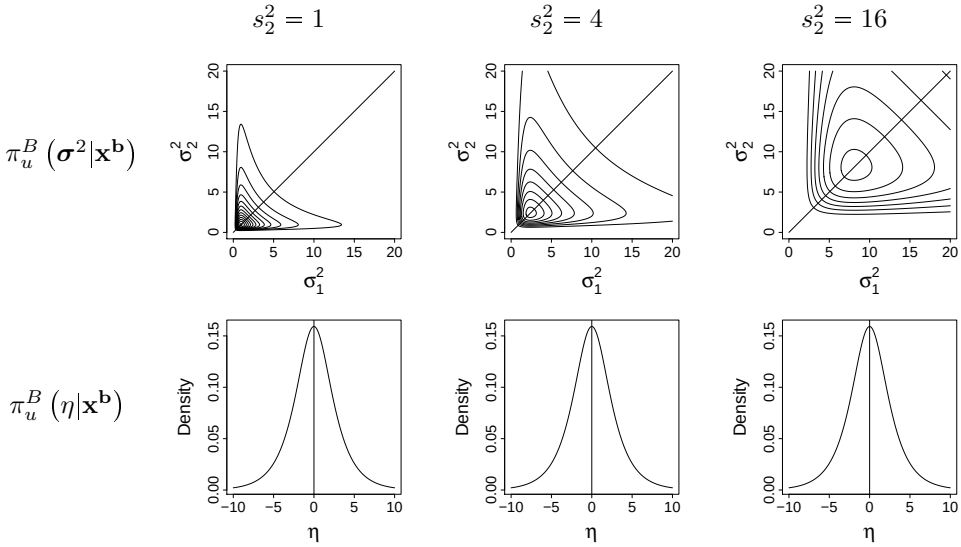


Figure 2.3: The marginal unconstrained BBF prior $\pi_u^B\left(\boldsymbol{\sigma}^2|\mathbf{x^b}\right)$ (top row) and the corresponding prior $\pi_u^B\left(\eta = \log\left(\sigma_1^2/\sigma_2^2\right)|\mathbf{x^b}\right)$ (bottom row) for sample variances $s_1^2 = 1$ and $s_2^2 \in \{1, 4, 16\}$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.075$. The prior $\pi_u^B\left(\eta|\mathbf{x^b}\right)$ is always balanced.

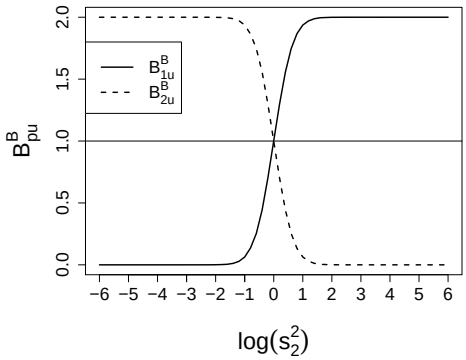


Figure 2.4: Bayes factors B_{1u}^B (solid line) and B_{2u}^B (dashed line) for sample variances $s_1^2 = 1$ and $s_2^2 \in [\exp(-6), \exp(6)]$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.075$. The Bayes factors favor the more parsimonious inequality constrained hypothesis if it is supported by the data. This shows that B_{1u}^B and B_{2u}^B function as Occam's razor.

(2.5). First, we use the unconstrained likelihood and Jeffreys prior. Second, in the denominator we integrate over an adjusted parameter space Ω_p^a, which will be defined shortly. We do not adjust the FBF marginal likelihoods under H_0 and H_u, that is, we set

$$m_0^{aF}(\mathbf{b}, \mathbf{x}) = m_0^F(\mathbf{b}, \mathbf{x}) \quad \text{and} \quad m_u^{aF}(\mathbf{b}, \mathbf{x}) = m_u^F(\mathbf{b}, \mathbf{x}). \tag{2.26}$$

The aFBF of H_p, $p = 1, 2$, against H_u is then given by

$$B_{pu}^{aF} = \frac{m_p^{aF}(\mathbf{b}, \mathbf{x})}{m_u^{aF}(\mathbf{b}, \mathbf{x})} = \frac{\int_{\Omega_p} \pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}\right) d\boldsymbol{\sigma}^2}{\int_{\Omega_p^a} \pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) d\boldsymbol{\sigma}^2} = \frac{P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}\right)}{P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p^a|\mathbf{x}^{\mathbf{b}}\right)}, \tag{2.27}$$

where $P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}\right)$ and $\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)$ are as in Equations (2.33) and (2.12), respectively. A derivation is given in Appendix 2.D.

Now, we want $P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p^a|\mathbf{x}^{\mathbf{b}}\right) = \int_{\Omega_p^a} \pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) d\boldsymbol{\sigma}^2 = 0.5$ (similar to $P^B\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}^{\mathbf{b}}\right)$ in Equation (2.19)) to ensure that the automatic Bayes factor B_{pu}^{aF} functions as Occam's razor when evaluating an inequality constrained hypothesis. To achieve this, we define the adjusted parameter space Ω_p^a, $p = 1, 2$, as

$$\Omega_1^a := \left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 < a\sigma_2^2\right\} \quad \text{and} \quad \Omega_2^a := \left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 > a\sigma_2^2\right\}, \tag{2.28}$$

where a is a constant chosen such that $P^F\left(\boldsymbol{\sigma}^2 \in \Omega_1^a|\mathbf{x}^{\mathbf{b}}\right) = P^F\left(\boldsymbol{\sigma}^2 \in \Omega_2^a|\mathbf{x}^{\mathbf{b}}\right) = 0.5$. Figure 2.5 illustrates this. The plot shows $\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)$ for sample variances $s_1^2 = 1$ and $s_2^2 = 4$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.1$. Two lines $\sigma_1^2 = a\sigma_2^2$ are depicted, one for $a = 1$ and one for $a = 0.25$. To determine Ω_1^a and Ω_2^a we proceed as follows. It can be seen that the probability mass in Ω_1 (i.e. above the line $\sigma_1^2 = 1 \cdot \sigma_2^2$) is larger than that in Ω_2. By tuning a we tilt the line $\sigma_1^2 = a\sigma_2^2$ such that the probability mass above and below the line is equal to 0.5. For the prior depicted in Figure 2.5 this is the case for $a = 0.25$. We thus have $\Omega_1^a = \left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 < 0.25 \cdot \sigma_2^2\right\}$ and $\Omega_2^a = \left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 > 0.25 \cdot \sigma_2^2\right\}$, and $P^F\left(\boldsymbol{\sigma}^2 \in \Omega_1^a|\mathbf{x}^{\mathbf{b}}\right) = P^F\left(\boldsymbol{\sigma}^2 \in \Omega_2^a|\mathbf{x}^{\mathbf{b}}\right) = 0.5$.

If we use $\mathbf{b} = (2/n_1, 2/n_2)'$ in order to satisfy the minimal information property, then it can be shown that $a = \frac{n_2(n_1-1)s_1^2}{n_1(n_2-1)s_2^2}$. In this case we can show that $P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p^a|\mathbf{x}^{\mathbf{b}}\right) = 0.5$ by transforming the integral

$$\begin{aligned}
P^F\left(\boldsymbol{\sigma}^2 \in \Omega_1^a|\mathbf{x}^{\mathbf{b}}\right) &= \int_{\Omega_1^a} \pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) d\boldsymbol{\sigma}^2 \\
&= \int_{\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 < a\sigma_2^2\}} \text{Inv-}\chi^2\left(\sigma_1^2|\nu_1, \tau_1^2\right) \text{Inv-}\chi^2\left(\sigma_2^2|\nu_2, \tau_2^2\right) d\boldsymbol{\sigma}^2 \\
&= \int_{\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 < \sigma_2^2\}} \text{Inv-}\chi^2\left(\sigma_1^2|1, \tau_1^2\right) \text{Inv-}\chi^2\left(\sigma_2^2|1, a\tau_2^2\right) d\boldsymbol{\sigma}^2 \\
&= \int_{\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 < \sigma_2^2\}} \text{Inv-}\chi^2\left(\sigma_1^2|1, \tau_1^2\right) \text{Inv-}\chi^2\left(\sigma_2^2|1, \tau_1^2\right) d\boldsymbol{\sigma}^2 \\
&= \int_{\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 < \sigma_2^2\}} \pi_u^{aF}\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) d\boldsymbol{\sigma}^2 = 0.5,
\end{aligned} \tag{2.29}$$

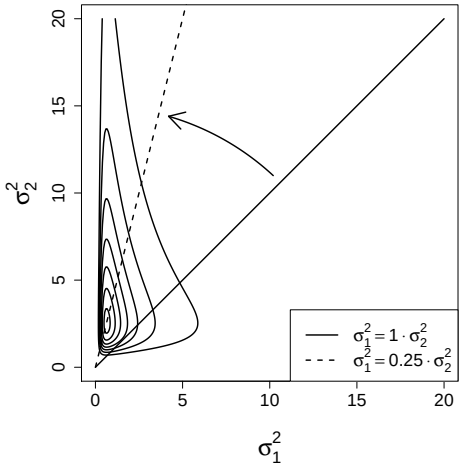


Figure 2.5: Marginal unconstrained FBF prior $\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)$ for sample variances $s_1^2 = 1$ and $s_2^2 = 4$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.1$. The probability mass above the line $\sigma_1^2 = a\sigma_2^2$, $a = 1$, is larger than that below it. We adjust the line by decreasing a until the probability mass above and below the line $\sigma_1^2 = a\sigma_2^2$ is equal to 0.5. For the depicted prior this is the case for $a = 0.25$.

with ν_j and τ_j^2, $j = 1, 2$, as in Equation (2.13). Here we used the result that if $\sigma^2 \sim \text{Inv-}\chi^2\left(\nu, \tau^2\right)$, then $a\sigma^2 \sim \text{Inv-}\chi^2\left(\nu, a\tau^2\right)$. The density

$$\pi_u^{aF}\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) = \text{Inv-}\chi^2\left(\sigma_1^2|1, \tau_1^2\right)\text{Inv-}\chi^2\left(\sigma_2^2|1, \tau_1^2\right) \tag{2.30}$$

can be regarded as the implicit unconstrained prior in the aFBF approach. Note that irrespective of the exact choice of $\mathbf{b}$ there always exists an a that yields $P^F\left(\boldsymbol{\sigma}^2 \in \Omega_1^a|\mathbf{x}^{\mathbf{b}}\right) = P^F\left(\boldsymbol{\sigma}^2 \in \Omega_2^a|\mathbf{x}^{\mathbf{b}}\right) = 0.5$.

Evaluation of the Method

We will now evaluate the aFBF according to the five properties discussed in Section 2.3:

1. *Proper priors.* As with the FBF, we must have $b_1 n_1 + b_2 n_2 > 2$ under H_0 and $b_j n_j > 1$, $j = 1, 2$, under H_1, H_2, and H_u to ensure that the priors are proper.

2. *Minimal information.* As was mentioned before, the minimal information property can be satisfied by setting $\mathbf{b} = (2/n_1, 2/n_2)'$.

3. *Scale invariance.* The aFBF is scale-invariant for the same reason that the FBF is (see Section 2.4.1).

4. *Balancedness.* In Equation (2.30) we have seen that the implicit unconstrained prior on $\boldsymbol{\sigma}^2$ is a product of two scaled inverse-χ^2 distributions with identical hyperparameters. Thus the corresponding prior on η is balanced (see Appendix 2.C).

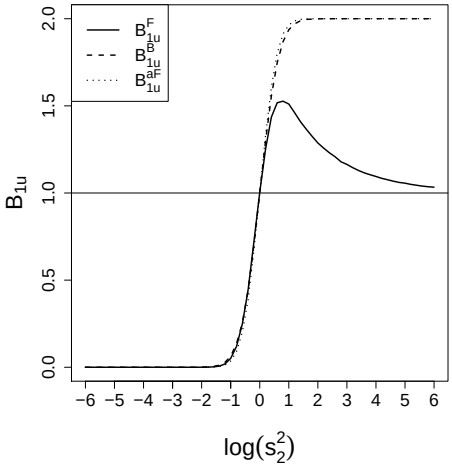


Figure 2.6: Bayes factors B_{1u}^F (solid line), B_{1u}^B (dashed line), and B_{1u}^{aF} (dotted line) for sample variances $s_1^2 = 1$ and $s_2^2 \in [\exp(-6), \exp(6)]$ and sample sizes $n_1 = n_2 = 20$. In the FBF and the aFBF the fractions are $b_1 = b_2 = 0.1$, while in the BBF we have $b_1 = b_2 = 0.075$. For $s_1^2 < s_2^2$ the Bayes factor B_{1u}^{aF} favors the more parsimonious inequality constrained hypothesis $H_1 \colon \sigma_1^2 < \sigma_2^2$. It thus functions as Occam's razor.

5. *Occam's razor.* Figure 2.6 shows the behavior of B_{1u}^{aF} (dotted line) as compared to B_{1u}^F (solid line) and B_{1u}^B (dashed line) for sample variances $s_1^2 = 1$ and $s_2^2 \in [\exp(-6), \exp(6)]$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.1$. For $s_1^2 < s_2^2$ the Bayes factor B_{1u}^{aF} favors the more parsimonious inequality constrained hypothesis $H_1 \colon \sigma_1^2 < \sigma_2^2$. It thus functions as Occam's razor.

2.5 Performance of the Bayes Factors

We present results of a simulation study investigating the performance of the three automatic Bayes factors. We consider two normal populations $X_1 \sim \mathrm{N}(0, 1)$ and $X_2 \sim \mathrm{N}(0, \sigma_2^2)$, where $\sigma_2^2 \in \{1.0, 1.5, 2.0, 2.5\}$. That is, we consider four effect sizes $\sigma_2^2/\sigma_1^2 \in \{1.0, 1.5, 2.0, 2.5\}$. A study by Ruscio and Roche (2012, Table 2) indicates that these population variance ratios roughly correspond to $\{\text{no}, \text{small}, \text{medium}, \text{large}\}$ effects in psychological research. We first investigate the strength of the evidence in favor of the true hypothesis H_t, $t = 0, 1$. The goal here is to see which automatic Bayes factor converges fastest to the true hypothesis. Following this, we consider frequentist error probabilities of selecting the wrong hypothesis. Note that from a Bayesian point of view these probabilities are of limited importance because Bayes factors are consistent in the sense that the evidence in favor of the true hypothesis grows to infinity as the sample size accumulates. These frequentist probabilities can be useful, however, to decide which automatic Bayes factor to use based on differences in error probability behavior.

2.5.1 Strength of Evidence in Favor of the True Hypothesis

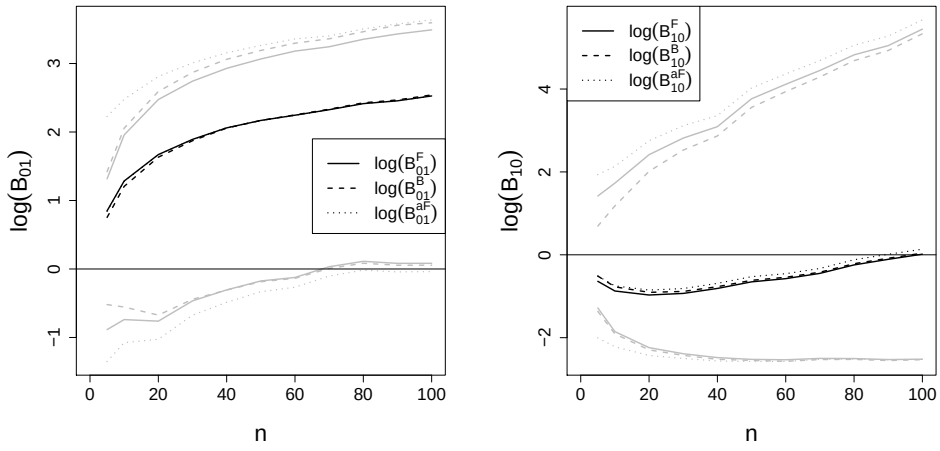
In this section we will investigate which automatic Bayes factor provides strongest evidence in favor of the true hypothesis. We shall use two measures of evidence. The first is the weight of evidence in favor of H_t against $H_{t'}$, where $t' = 1$ if $t = 0$ and $t' = 0$ otherwise. The weight of evidence is given by the logarithm of the Bayes factor, that is, $\log(B_{tt'})$. The second measure of evidence we use is the posterior probability of the true hypothesis. Assuming that all hypotheses are equally likely a priori (i.e. $P(H_0) = P(H_1) = P(H_2) = 1/3$, which is a standard default choice), it is given by $P(H_t|\mathbf{x}) = \frac{m_t(\mathbf{b},\mathbf{x})}{m_0(\mathbf{b},\mathbf{x})+m_1(\mathbf{b},\mathbf{x})+m_2(\mathbf{b},\mathbf{x})}$, where $m_t(\mathbf{b},\mathbf{x})$ denotes the marginal likelihood under H_t. Both measures of evidence are computed for the FBF, the BBF, and the aFBF.

We drew 5000 samples of size $n_1 = n_2 = n \in \{5, 10, 20, \ldots, 100\}$ from X_1 and X_2. Denote these samples by $\mathbf{x}^{(m)} = \left(\mathbf{x}_1^{(m)}, \mathbf{x}_2^{(m)}\right)'$, $m = 1, \ldots, 5000$. For each $\mathbf{x}^{(m)}$ we computed the two measures of evidence $\log(B_{tt'})^{(m)}$ and $P\left(H_t|\mathbf{x}^{(m)}\right)$. Eventually, we computed the median of $\left\{\log(B_{tt'})^{(m)}\right\}_{m=1}^{5000}$ and $\left\{P\left(H_t|\mathbf{x}^{(m)}\right)\right\}_{m=1}^{5000}$ to estimate the average evidence in favor of H_t, as well as the 2.5%- and 97.5%-quantile to obtain an indication of the variability of the evidence.

Figure 2.7 shows the results for the weight of evidence, $\log(B_{tt'})$. The plots show the median (black lines) and the 2.5%- and 97.5%-quantile (gray lines) as a function of the common sample size n for each $\sigma_2^2 \in \{1.0, 1.5, 2.0, 2.5\}$. It can be seen that the three automatic Bayes factors provide similarly strong median evidence in favor of the true hypothesis (panels (a) to (d)). In panel (a) the dotted line for the aFBF is actually covered by the lines for the FBF and the BBF. If there is a positive effect (panels (b) to (d)), then the aFBF provides slightly stronger evidence in favor of the true hypothesis H_1 than the FBF and the BBF (as can be seen from the lines for the median and the 97.5%-quantile). The BBF, on the other hand, provides somewhat weaker evidence in favor of H_1. This is because the balanced prior slightly shrinks the posterior towards $\sigma_1^2 = \sigma_2^2$, which results in a loss of evidence in favor of an inequality constrained hypothesis that is supported by the data. The FBF and the aFBF are not affected by such shrinkage. Figure 2.8 shows the simulation results for the posterior probability of the true hypothesis, $P(H_t|\mathbf{x})$. In the legends the superscripts F, B, and aF denote on which Bayes factor the posterior probability is based. The results are in line with those from Figure 2.7. In fact, the advantage of the aFBF over the FBF and the BBF in terms of strength of evidence is a bit more pronounced. Overall, it can be concluded that the aFBF performs best: under H_0 it performs about as good as the FBF and the BBF, while under H_1 it slightly outperforms the latter two.

2.5.2 Frequentist Error Probabilities

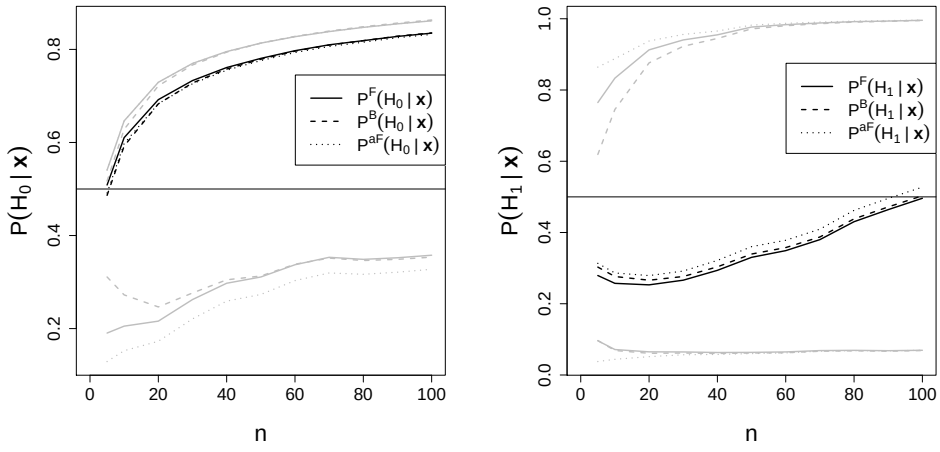
Table 2.1 shows simulated frequentist error probabilities of the three automatic Bayes factors and the likelihood-ratio (LR) test for $\sigma_1^2 = 1$ and $\sigma_2^2 \in \{1.0, 1.5, 2.0, 2.5\}$. For each σ_2^2 we drew 5000 samples of size $n_1 = n_2 = n \in \{5, 50, 500\}$ from $X_1 \sim \mathrm{N}(0, 1)$ and $X_2 \sim \mathrm{N}(0, \sigma_2^2)$. On each sample we computed the Bayes factors and the LR test. In the Bayesian testing approach an error occurs if the true hypothesis H_t does not have the largest posterior probability, that is, if $P\left(H_{t'}|\mathbf{x}^{(m)}\right) > P\left(H_t|\mathbf{x}^{(m)}\right)$ for



(a) $\sigma_2^2 = 1.0$ (b) $\sigma_2^2 = 1.5$

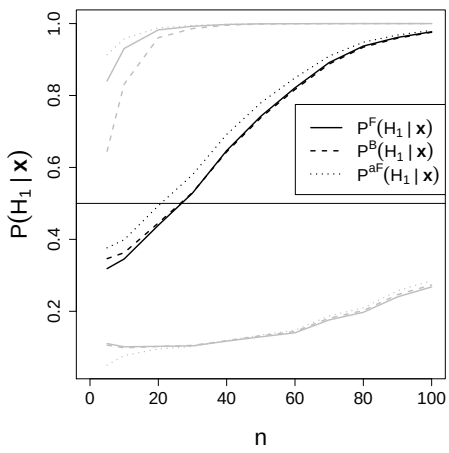
(c) $\sigma_2^2 = 2.0$ (d) $\sigma_2^2 = 2.5$

Figure 2.7: Results of a simulation study investigating the performance of the FBF, the BBF, and the aFBF in testing variances of two normal populations $X_1 \sim \mathrm{N}(0, 1)$ and $X_2 \sim \mathrm{N}(0, \sigma_2^2)$, where $\sigma_2^2 \in \{1.0, 1.5, 2.0, 2.5\}$. The black lines depict the median weight of evidence in favor of the true hypothesis H_t, $\log(B_{tt'})$, as a function of the common sample size $n_1 = n_2 = n$. The gray lines depict the 2.5%- and 97.5%-quantile. It can be seen that if there is a positive effect (i.e. if $\sigma_1^2 < \sigma_2^2$), then the aFBF provides strongest evidence in favor of the true hypothesis H_1.

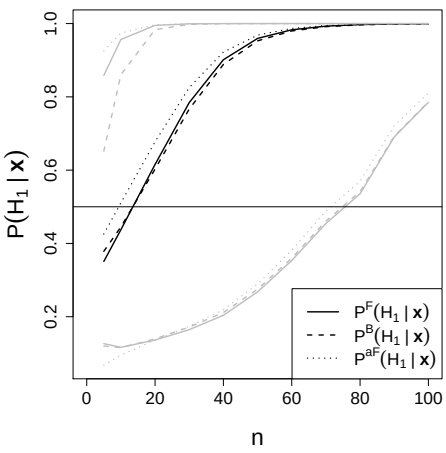


(a) $\sigma_2^2 = 1.0$

(b) $\sigma_2^2 = 1.5$



(c) $\sigma_2^2 = 2.0$



(d) $\sigma_2^2 = 2.5$

Figure 2.8: Results of a simulation study investigating the performance of the FBF, the BBF, and the aFBF in testing variances of two normal populations $X_1 \sim \mathrm{N}(0, 1)$ and $X_2 \sim \mathrm{N}(0, \sigma_2^2)$, where $\sigma_2^2 \in \{1.0, 1.5, 2.0, 2.5\}$. The black lines depict the median posterior probability of the true hypothesis H_t, $P\left(H_t|\mathbf{x}\right)$, as a function of the common sample size $n_1 = n_2 = n$. The gray lines depict the 2.5%- and 97.5%-quantile. In the legends the superscripts F, B, and aF denote on which Bayes factor the posterior probability is based. It can be seen that if there is a positive effect (i.e. if $\sigma_1^2 < \sigma_2^2$), then the aFBF provides strongest evidence in favor of the true hypothesis H_1.

Table 2.1: Frequentist error probabilities of the three automatic Bayes factors and the likelihood-ratio (LR) test for $\sigma_1^2 = 1$, $\sigma_2^2 \in \{1.0, 1.5, 2.0, 2.5\}$, and $n_1 = n_2 = n \in \{5, 50, 500\}$. In the LR test we set $\alpha = 0.05$. It can be seen that under H_1 the aFBF has lower error probabilities than the FBF and the BBF.

σ_2^2	1.0			1.5			2.0			2.5		
n	5	50	500	5	50	500	5	50	500	5	50	500
FBF	0.23	0.07	0.02	0.80	0.66	0.01	0.72	0.28	0.00	0.65	0.09	0.00
BBF	0.26	0.07	0.02	0.79	0.66	0.01	0.69	0.28	0.00	0.62	0.09	0.00
aFBF	0.36	0.08	0.02	0.72	0.63	0.01	0.60	0.26	0.00	0.54	0.08	0.00
LR test	0.05	0.05	0.05	0.94	0.71	0.00	0.92	0.33	0.00	0.89	0.11	0.00

some $t' \neq t$. Here again we assumed equal prior probabilities of the hypotheses. In the frequentist approach an error occurs under H_0 if $p < \alpha$ and under H_1 if $p > \alpha \vee \left(p < \alpha \wedge s_1^2 > s_2^2\right)$. In the present simulation we set $\alpha = 0.05$. Table 2.1 shows the proportions of errors in the 5000 samples. It can be seen that the error probabilities of the three automatic Bayes factors are quite similar. Under H_0 the aFBF shows somewhat larger error probabilities. Under H_1, however, it has lower error probabilities than the FBF and the BBF, particularly for $n = 5$. Moreover, it can be seen that under H_1 the Bayes factors have lower error probabilities than the LR test. While the differences are considerable for $n = 5$, the LR test closes the gap as the sample size increases. One final remark concerns the error probabilities under H_0: While the LR test has unconditional error probabilities equal to $\alpha = 0.05$ regardless of the sample size, the conditional error probabilities of the three Bayes factors decrease as the sample size increases. This illustrates that the automatic Bayes factors are consistent whereas the p-value is not.

Additional insight into the performance of the three automatic Bayes factors is given in Table 2.2. It is well-known that p-values tend to overstate the evidence against the null hypothesis and that methods based on comparing likelihoods (such as Bayes factors and posterior probabilities of hypotheses) commonly yield weaker evidence against the null (see, for example, Berger & Sellke, 1987; Held, 2010; Sellke, Bayarri, & Berger, 2001). Table 2.2 shows that this also holds for the three automatic Bayes factors discussed in this chapter. The table can be read as follows. For sample sizes of $n_1 = n_2 = n = 5$ and sample variances of $s_1^2 = 1$ and $s_2^2 = 9.60$, the standard likelihood-ratio test of equality of variances yields a two-sided p-value of 0.05. The posterior probabilities of H_0 based on these sample data are $P^F(H_0|\mathbf{x}) = 0.26$, $P^B(H_0|\mathbf{x}) = 0.34$, and $P^{aF}(H_0|\mathbf{x}) = 0.19$. From the frequentist significance test we would thus conclude that there is evidence against H_0, whereas the posterior probabilities tell us that there is some evidence for H_0 given the observed data. This discrepancy between the p-value and the posterior probabilities of H_0 becomes even more pronounced for larger sample sizes. A similar picture emerges for $p = 0.01$: While the p-value tells us that there is strong evidence against H_0, it is difficult to rule out H_0 given posterior probabilities roughly between 0.1 and 0.3. It can be seen that the posterior probabilities of H_0 decrease as the p-value decreases. This suggests that only very small p-values should be considered indicative of evidence against H_0, particularly if sample sizes are large.

Table 2.2: Comparison of two-sided p-values and posterior probabilities of H_0, denoted by $P(H_0|\mathbf{x})$. The superscripts F, B, and aF denote on which Bayes factor $P(H_0|\mathbf{x})$ is based. For example, sample sizes of $n_1 = n_2 = n = 5$ and sample variances of $s_1^2 = 1.00$ and $s_2^2 = 9.60$ yield a p-value of 0.05 and posterior probabilities of H_0 of 0.26, 0.34, and 0.19. It can be seen that while the p-values indicate evidence against H_0, the posterior probabilities tell us that H_0 is quite likely given the sample data.

		$p = 0.05$				$p = 0.01$			
n	s_1^2	s_2^2	$P^F(H_0\|\mathbf{x})$	$P^B(H_0\|\mathbf{x})$	$P^{aF}(H_0\|\mathbf{x})$	s_2^2	$P^F(H_0\|\mathbf{x})$	$P^B(H_0\|\mathbf{x})$	$P^{aF}(H_0\|\mathbf{x})$
5	1.00	9.60	0.26	0.34	0.19	23.15	0.11	0.28	0.07
10	1.00	4.03	0.29	0.34	0.23	6.54	0.11	0.20	0.08
20	1.00	2.53	0.34	0.36	0.29	3.43	0.13	0.16	0.10
50	1.00	1.76	0.43	0.43	0.39	2.11	0.17	0.18	0.14
100	1.00	1.49	0.51	0.50	0.48	1.69	0.21	0.21	0.19

2.6 Empirical Data Examples

In this section we apply the three automatic Bayes factors to two empirical data sets.

2.6.1 Example 1: Variability of Intelligence in Children (Arden & Plomin, 2006)

We first consider a study by Arden and Plomin (2006) investigating differences in variance of intelligence between girls and boys. Psychological research has consistently found males to be more variable in intellectual abilities than females (e.g. Feingold, 1992). Arden and Plomin therefore assumed that this finding would also apply to children. Their dependent variable of interest was a general ability factor extracted from several tests of verbal and non-verbal ability. The authors expected that boys would show larger variance on this factor than girls, which can be formulated in the hypothesis $H_1\colon \sigma_f^2 < \sigma_m^2$, where σ_f^2 and σ_m^2 denote the population variances of females and males, respectively. The competing hypotheses are $H_0\colon \sigma_f^2 = \sigma_m^2$ and $H_2\colon \sigma_f^2 > \sigma_m^2$.

In samples of $n_f = 1366$ girls and $n_m = 1136$ boys of age 10, Arden and Plomin found sample variances of $s_f^2 = 0.92$ and $s_m^2 = 1.10$. Table 2.3 provides the Bayes factors B_{10} and B_{12} and the posterior probabilities of H_0, H_1, and H_2 (assuming equal prior probabilities) for these sample data. As can be seen, the posterior probabilities of H_0, H_1, and H_2 are approximately 0.13, 0.87, and 0.00 for all three automatic Bayes factors. An immediate conclusion we can draw from these results is that we can basically rule out H_2. The Bayes factors B_{10} and B_{12}, and the posterior probability of H_1, $P(H_1|\mathbf{x})$, indicate positive evidence in favor of H_1. However, the evidence does not appear to be strong enough to completely rule out H_0. The two-sided p-value for these data obtained from the standard likelihood-ratio test equals 0.002, which would commonly be interpreted as sufficient evidence to reject H_0 in favor of the two-sided alternative.

Table 2.3: Results for two empirical data examples.

	Example 1					Example 2				
	B_{10}	B_{12}	$P(H_0\|\mathbf{x})$	$P(H_1\|\mathbf{x})$	$P(H_2\|\mathbf{x})$	B_{01}	B_{02}	$P(H_0\|\mathbf{x})$	$P(H_1\|\mathbf{x})$	$P(H_2\|\mathbf{x})$
FBF	6.32	1176.58	0.14	0.86	0.00	7.14	5.52	0.76	0.10	0.14
BBF	6.43	1261.63	0.13	0.87	0.00	7.73	4.96	0.75	0.10	0.15
aFBF	6.68	1316.52	0.13	0.87	0.00	7.21	5.47	0.76	0.10	0.14

2.6.2 Example 2: Precision of Burn Wound Assessments (N. A. J. Martin, Lundy, & Rickard, 2014)

We next reanalyze data from a study by Martin et al. (2014) investigating the precision of burn wound assessments by UK Armed Forces medical personnel. The percentage of the total body surface area that is burned (%TBSA burned) is a very important measure in the treatment of burn victims. The authors had two groups of medical personnel estimate the %TBSA burned for one particular burn case. The first group consisted of $n_1 = 20$ experienced burn specialists, while the second group consisted of $n_2 = 40$ relatively inexperienced participants of a surgical training course. Martin et al. expected the experienced burn specialists to be less variable in their %TBSA burned estimates than the inexperienced medical personnel. This expectation can be formulated in the hypothesis $H_1 : \sigma_1^2 < \sigma_2^2$, the competing hypotheses being $H_0 : \sigma_1^2 = \sigma_2^2$ and $H_2 : \sigma_1^2 > \sigma_2^2$.

Martin et al. found sample variances of $s_1^2 = 105.88$ and $s_2^2 = 100.60$. The two-sided p-value obtained from the standard likelihood-ratio test equals $p = 0.86$ for these sample data. From this p-value it can be concluded that there is not enough evidence to reject the null hypothesis that the two groups are equally heterogeneous. However, we cannot conclude that there is evidence in favor of the null hypothesis since p-values do not imply this kind of information. The p-value of 0.86 thus leaves us in a state of ignorance. The Bayes factor on the other hand can be used to quantify the relative evidence in favor of a null hypothesis. Table 2.3 provides the Bayes factors B_{01} and B_{02} and the posterior probabilities of H_0, H_1, and H_2 (assuming equal prior probabilities). The Bayes factors and the posterior probability of H_0, $P(H_0|\mathbf{x})$, indicate positive evidence in favor of H_0. In particular, the posterior probability of H_0 is approximately 0.76 for all three automatic Bayes factors. However, the posterior probabilities of H_1 and H_2 are between 0.10 and 0.15, indicating that it is difficult to completely rule out either of the two hypotheses based on the sample data.

2.7 Discussion

In this chapter we presented three automatic Bayes factors for testing variances of two independent normal distributions: the FBF, the BBF, and the aFBF. The three Bayes factors are fully automatic and thus readily applicable. All the user needs to provide is the two sample sizes and sample variances. This makes the Bayes factors particularly valuable for both statisticians and applied researchers who are interested in a user-friendly Bayesian method for testing two variances.

The methods were theoretically evaluated on the basis of five properties: proper

priors, minimal information, scale invariance, balancedness, and Occam's razor. As was shown, the FBF satisfies neither the balancedness property nor the Occam's razor property when testing inequality constraints on variances. The BBF and the aFBF, on the other hand, satisfy all five properties. In the BBF, an automatic balanced prior is constructed based on equal prior distributions for the variances with minimal information. In the aFBF the FBF marginal likelihood is adjusted such that it adequately incorporates the parsimony of an inequality constrained hypothesis. The simulation study indicates that the aFBF provides strongest evidence in favor of a true inequality constrained hypothesis. The slightly worse performance of the BBF is caused by the fact that the balanced prior shrinks the posterior towards the boundary of the constrained space where $\sigma_1^2 = \sigma_2^2$, resulting in weaker evidence in favor of a true inequality constrained hypothesis. The FBF and the aFBF, on the other hand, are not affected by prior shrinkage.

One possible point of debate relating to all three Bayes factors is the the choice of the fraction b. In this chapter we used the minimal information approach to specifying b, which is a widely accepted principle. However, there are other approaches to specifying b, each of which pursuing a different goal (see, for example, Conigliani & O'Hagan, 2000; O'Hagan, 1995). All important formulas in this chapter are expressed in terms of b. It is therefore straightforward to use a different b if desired.

There are two natural extensions of our approach to testing two variances. First, we are often interested in testing variances of $J > 2$ independent populations. Relevant hypotheses in this case include the null hypothesis $H_0\colon \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_J^2$, the order constrained hypothesis $H_1\colon \sigma_1^2 < \sigma_2^2 < \cdots < \sigma_J^2$, and hypotheses with combinations of equality and inequality constraints, for example $H_2\colon \sigma_1^2 = \sigma_2^2 < \cdots < \sigma_J^2$. The second natural extension is testing variances of dependent populations, which is relevant when analyzing repeated measurement data. Based on our findings the automatic Bayes factors discussed in this chapter will prove useful for these more complex testing problems.

2.A Derivation of $m_0^F(\mathbf{b}, \mathbf{x})$

Plugging $f_0\left(\mathbf{x}|\boldsymbol{\mu}, \sigma^2\right)$, $f_0\left(\mathbf{x}|\boldsymbol{\mu}, \sigma^2\right)^{\mathbf{b}}$, and $\pi_0^N\left(\boldsymbol{\mu}, \sigma^2\right)$ into Equation (2.5) gives us

$$m_0^F(\mathbf{b}, \mathbf{x}) = \frac{\int_{\Omega_0}\int_{\mathbb{R}^2} f_0\left(\mathbf{x}|\boldsymbol{\mu}, \sigma^2\right)\pi_0^N\left(\boldsymbol{\mu}, \sigma^2\right)d\boldsymbol{\mu}d\sigma^2}{\int_{\Omega_0}\int_{\mathbb{R}^2} f_0\left(\mathbf{x}|\boldsymbol{\mu}, \sigma^2\right)^{\mathbf{b}}\pi_0^N\left(\boldsymbol{\mu}, \sigma^2\right)d\boldsymbol{\mu}d\sigma^2} = \frac{m_0^F(\mathbf{x})}{m_0^F\left(\mathbf{x}^{\mathbf{b}}\right)}. \tag{2.31}$$

We first derive the denominator $m_0^F\left(\mathbf{x}^{\mathbf{b}}\right)$. Note that the Jeffreys prior can be written as $\pi_0^N\left(\boldsymbol{\mu}, \sigma^2\right) = k_0\sigma^{-2}$, where k_0 is an unspecified normalizing constant (see, for

instance, O'Hagan, 1995). Then

$$\begin{aligned}
m_0^F\left(\mathbf{x}^{\mathbf{b}}\right) &= \int_{\mathbb{R}^+}\int_{\mathbb{R}^2} f\left(\mathbf{x}_1|\mu_1,\sigma^2\right)^{b_1} f\left(\mathbf{x}_2|\mu_2,\sigma^2\right)^{b_2} k_0\sigma^{-2}d\boldsymbol{\mu}d\sigma^2 \\
&= k_0\int_{\mathbb{R}^+}\int_{\mathbb{R}} f\left(\mathbf{x}_1|\mu_1,\sigma^2\right)^{b_1} d\mu_1 \int_{\mathbb{R}} f\left(\mathbf{x}_2|\mu_2,\sigma^2\right)^{b_2} d\mu_2\, \sigma^{-2}d\sigma^2 \\
&= k_0\int_{\mathbb{R}^+} (b_1n_1)^{-\frac{1}{2}}\left(\sigma^2 2\pi\right)^{-\frac{b_1n_1-1}{2}} \exp\left(-\frac{b_1\left(n_1-1\right)s_1^2}{2\sigma^2}\right) \\
&\qquad (b_2n_2)^{-\frac{1}{2}}\left(\sigma^2 2\pi\right)^{-\frac{b_2n_2-1}{2}} \exp\left(-\frac{b_2\left(n_2-1\right)s_2^2}{2\sigma^2}\right)\sigma^{-2}d\sigma^2 \\
&= k_0\left(b_1b_2\right)^{-\frac{1}{2}}\left(n_1n_2\right)^{-\frac{1}{2}}(2\pi)^{-\frac{b_1n_1+b_2n_2-2}{2}} \\
&\qquad \int_{\mathbb{R}^+}\left(\sigma^2\right)^{-\left(\frac{b_1n_1+b_2n_2-2}{2}+1\right)} \exp\left(-\frac{b_1\left(n_1-1\right)s_1^2+b_2\left(n_2-1\right)s_2^2}{2\sigma^2}\right)d\sigma^2 \\
&= k_0\left(b_1b_2\right)^{-\frac{1}{2}}\left(n_1n_2\right)^{-\frac{1}{2}}\pi^{-\frac{b_1n_1+b_2n_2-2}{2}}\Gamma\left(\frac{b_1n_1+b_2n_2-2}{2}\right) \\
&\qquad \left(b_1\left(n_1-1\right)s_1^2+b_2\left(n_2-1\right)s_2^2\right)^{-\frac{b_1n_1+b_2n_2-2}{2}},
\end{aligned} \tag{2.32}$$

where the integrand in the last but one line is the kernel of a scaled inverse-χ^2 distribution with parameters $\nu = b_1n_1 + b_2n_2 - 2$ and $\tau^2 = \frac{b_1(n_1-1)s_1^2+b_2(n_2-1)s_2^2}{b_1n_1+b_2n_2-2}$. We obtain $m_0^F(\mathbf{x})$ by setting $b_1 = b_2 = 1$ in the expression for $m_0^F\left(\mathbf{x}^{\mathbf{b}}\right)$. Dividing $m_0^F(\mathbf{x})$ by $m_0^F\left(\mathbf{x}^{\mathbf{b}}\right)$ eventually yields the expression given in Equation (2.8). Note that the unspecified constant k_0 cancels out in this step.

2.B Probability That $\boldsymbol{\sigma}^2$ Is in Ω_p

For the FBF we have

$$P^F\left(\boldsymbol{\sigma}^2\in\Omega_p|\mathbf{x}\right) = \int_{\Omega_p}\text{Inv-}\chi^2\left(\sigma_1^2|n_1-1,s_1^2\right)\text{Inv-}\chi^2\left(\sigma_2^2|n_2-1,s_2^2\right)d\boldsymbol{\sigma}^2, \quad p=1,2, \tag{2.33}$$

and

$$\begin{aligned}
P^F\left(\boldsymbol{\sigma}^2\in\Omega_p|\mathbf{x}^{\mathbf{b}}\right) = \int_{\Omega_p}&\text{Inv-}\chi^2\left(\sigma_1^2|b_1n_1-1,\frac{b_1\left(n_1-1\right)s_1^2}{b_1n_1-1}\right) \\
&\text{Inv-}\chi^2\left(\sigma_2^2|b_2n_2-1,\frac{b_2\left(n_2-1\right)s_2^2}{b_2n_2-1}\right)d\boldsymbol{\sigma}^2, \quad p=1,2.
\end{aligned} \tag{2.34}$$

For the BBF we have

$$\begin{aligned}
P^B\left(\boldsymbol{\sigma}^2\in\Omega_p|\mathbf{x}\right) = \int_{\Omega_p}&\text{Inv-}\chi^2\left(\sigma_1^2|n_1+\nu_\bullet-1,\frac{\left(n_1-1\right)s_1^2+\nu_\bullet\tau_\bullet^2}{n_1+\nu_\bullet-1}\right) \\
&\text{Inv-}\chi^2\left(\sigma_2^2|n_2+\nu_\bullet-1,\frac{\left(n_2-1\right)s_2^2+\nu_\bullet\tau_\bullet^2}{n_2+\nu_\bullet-1}\right)d\boldsymbol{\sigma}^2, \quad p=1,2.
\end{aligned} \tag{2.35}$$

The integrals cannot be solved analytically, but they can be approximated numerically using Monte Carlo methods: first we draw samples from the two scaled inverse-χ^2 distributions Inv-$\chi^2\left(\sigma_1^2|\nu_1,\tau_1^2\right)$ and Inv-$\chi^2\left(\sigma_2^2|\nu_2,\tau_2^2\right)$. An approximation of the integral is then given by the proportion of draws that fall in Ω_p.

2.C Distribution of $\eta = \log\left(\sigma_1^2/\sigma_2^2\right)$

Let $\pi_{\sigma_1^2,\sigma_2^2}\left(\sigma_1^2,\sigma_2^2\right) = \text{Inv-}\chi^2_{\sigma_1^2}\left(\sigma_1^2|\nu_1,\tau_1^2\right)\text{Inv-}\chi^2_{\sigma_2^2}\left(\sigma_2^2|\nu_2,\tau_2^2\right)$ be the joint distribution of σ_1^2 and σ_2^2. We derive the distribution of $\eta = \log\left(\sigma_1^2/\sigma_2^2\right)$. To do so we first derive the distribution of $\zeta = \sigma_1^2/\sigma_2^2$ and subsequently apply the transformation $\eta = \log(\zeta)$. To determine the distribution of ζ we first determine the joint distribution of ζ and σ_2^2, which is given by

$$\begin{aligned}\pi_{\zeta,\sigma_2^2}\left(\zeta,\sigma_2^2\right) &= \left|\frac{\partial\sigma_1^2}{\partial\zeta}\right| \pi_{\sigma_1^2,\sigma_2^2}\left(\sigma_1^2,\sigma_2^2\right) \\ &= \frac{\left(\frac{\nu_1\tau_1^2}{2}\right)^{\frac{\nu_1}{2}}\left(\frac{\nu_2\tau_2^2}{2}\right)^{\frac{\nu_2}{2}}}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\zeta^{-\left(\frac{\nu_1}{2}+1\right)}\left(\sigma_2^2\right)^{-\left(\frac{\nu_1+\nu_2}{2}+1\right)}\exp\left(-\frac{\nu_1\tau_1^2/\zeta+\nu_2\tau_2^2}{2\sigma_2^2}\right).\end{aligned} \tag{2.36}$$

We then obtain the marginal distribution of ζ by integrating out σ_2^2:

$$\begin{aligned}\pi_\zeta(\zeta) &= \int_{\mathbb{R}^+}\pi_{\zeta,\sigma_2^2}\left(\zeta,\sigma_2^2\right)d\sigma_2^2 \\ &= \frac{\left(\nu_1\tau_1^2\right)^{\frac{\nu_1}{2}}\left(\nu_2\tau_2^2\right)^{\frac{\nu_2}{2}}\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\zeta^{-\left(\frac{\nu_1}{2}+1\right)}\left(\nu_1\tau_1^2\zeta^{-1}+\nu_2\tau_2^2\right)^{-\frac{\nu_1+\nu_2}{2}}.\end{aligned} \tag{2.37}$$

Eventually, we obtain the distribution of η by applying the transformation $\eta = \log(\zeta)$:

$$\begin{aligned}\pi_\eta(\eta) &= \left|\frac{d\exp(\eta)}{d\eta}\right| \pi_\zeta\left(\exp(\eta)\right) \\ &= \frac{\left(\nu_1\tau_1^2\right)^{\frac{\nu_1}{2}}\left(\nu_2\tau_2^2\right)^{\frac{\nu_2}{2}}\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\exp(-\eta)^{\frac{\nu_1}{2}}\left(\nu_1\tau_1^2\exp(-\eta)+\nu_2\tau_2^2\right)^{-\frac{\nu_1+\nu_2}{2}}.\end{aligned} \tag{2.38}$$

Next, we show that $\pi_\eta(\eta)$ is balanced if and only if $\pi_{\sigma_1^2,\sigma_2^2}\left(\sigma_1^2,\sigma_2^2\right) = \text{Inv-}\chi^2_{\sigma_1^2}\left(\sigma_1^2|\nu,\tau^2\right)\text{Inv-}\chi^2_{\sigma_2^2}\left(\sigma_2^2|\nu,\tau^2\right)$. In other words, we show that $\pi_\eta(\eta)$ is symmetric about 0 and nonincreasing in $|\eta|$ if and only if $\nu_1 = \nu_2 \wedge \tau_1^2 = \tau_2^2$. We first use a proof by contrapositive to show that if $\pi_\eta(\eta)$ is balanced, then $\nu_1 = \nu_2 \wedge \tau_1^2 = \tau_2^2$. Assume that $\neg\left(\nu_1 = \nu_2 \wedge \tau_1^2 = \tau_2^2\right)$. We show that $\pi_\eta(\eta)$ is not balanced. We consider three cases.

Case 1. $\nu_1 \neq \nu_2 \wedge \tau_1^2 \neq \tau_2^2$. Note that if $\pi_\eta(\eta)$ was symmetric about 0 and nonincreasing in $|\eta|$, then it would have a single mode at 0. In this case we would have $\frac{d}{d\eta}\pi_\eta(0) = 0$. However, if $\nu_1 \neq \nu_2 \wedge \tau_1^2 \neq \tau_2^2$, then $\frac{d}{d\eta}\pi_\eta(0) \neq 0$, which shows that $\pi_\eta(\eta)$ is not balanced.

Case 2. $\nu_1 = \nu_2 \wedge \tau_1^2 \neq \tau_2^2$. Note that if $\pi_\eta(\eta)$ was symmetric about 0, then we would have $\pi_\eta(\eta) = \pi_\eta(-\eta)$. Let $\nu_1 = \nu_2 = \nu$. Then

$$\begin{aligned}\pi_\eta(\eta) &= \frac{\Gamma(\nu)}{\Gamma\left(\frac{\nu}{2}\right)^2}\left(\tau_1^2\tau_2^2\right)^{\frac{\nu}{2}}\left(\tau_1^2\exp\left(-\frac{\eta}{2}\right)+\tau_2^2\exp\left(\frac{\eta}{2}\right)\right)^{-\nu} \\ &\neq \frac{\Gamma(\nu)}{\Gamma\left(\frac{\nu}{2}\right)^2}\left(\tau_1^2\tau_2^2\right)^{\frac{\nu}{2}}\left(\tau_1^2\exp\left(\frac{\eta}{2}\right)+\tau_2^2\exp\left(-\frac{\eta}{2}\right)\right)^{-\nu} = \pi_\eta(-\eta),\end{aligned} \tag{2.39}$$

since $\tau_1^2 \neq \tau_2^2$. This shows that $\pi_\eta(\eta)$ is not symmetric about 0, and thus not balanced.

Case 3. $\nu_1 \neq \nu_2 \wedge \tau_1^2 = \tau_2^2$. The argument is analogous to that in Case 2.

We next show that if $\nu_1 = \nu_2 \wedge \tau_1^2 = \tau_2^2$, then $\pi_\eta(\eta)$ is balanced. Let $\nu_1 = \nu_2 = \nu$ and $\tau_1^2 = \tau_2^2 = \tau^2$. Then $\pi_\eta(\eta)$ is symmetric about 0 since

$$\begin{aligned}\pi_\eta(\eta) &= \frac{\Gamma(\nu)}{\Gamma\left(\frac{\nu}{2}\right)^2}\left(\exp\left(-\frac{\eta}{2}\right)+\exp\left(\frac{\eta}{2}\right)\right)^{-\nu} \\ &= \frac{\Gamma(\nu)}{\Gamma\left(\frac{\nu}{2}\right)^2}\left(\exp\left(\frac{\eta}{2}\right)+\exp\left(-\frac{\eta}{2}\right)\right)^{-\nu} = \pi_\eta(-\eta).\end{aligned} \tag{2.40}$$

Eventually, note that $\exp(-\eta/2) + \exp(\eta/2)$ is strictly monotonically increasing for $|\eta| = \eta > 0$, in which case the inverse of this expression and $\pi_\eta(\eta)$ are strictly monotonically decreasing (i.e. nonincreasing).

2.D Derivation of B_{pu}^{aF}

$$\begin{aligned}B_{pu}^{aF} &= \frac{m_p^{aF}(\mathbf{b},\mathbf{x})}{m_u^{aF}(\mathbf{b},\mathbf{x})} = \frac{m_p^{aF}(\mathbf{b},\mathbf{x})}{m_u^{F}(\mathbf{b},\mathbf{x})} \\ &= \frac{\frac{\int_{\Omega_p}\int_{\mathbb{R}^2} f_u\left(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)\pi_u^N\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}{\int_{\Omega_p^a}\int_{\mathbb{R}^2} f_u(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2)^{\mathbf{b}}\pi_u^N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}}{\frac{\int_{\Omega_u}\int_{\mathbb{R}^2} f_u(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2)\pi_u^N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}{\int_{\Omega_u}\int_{\mathbb{R}^2} f_u(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2)^{\mathbf{b}}\pi_u^N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}} = \frac{\frac{\int_{\Omega_p}\int_{\mathbb{R}^2} f_u\left(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)\pi_u^N\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}{\int_{\Omega_u}\int_{\mathbb{R}^2} f_u(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2)\pi_u^N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}}{\frac{\int_{\Omega_p^a}\int_{\mathbb{R}^2} f_u(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2)^{\mathbf{b}}\pi_u^N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}{\int_{\Omega_u}\int_{\mathbb{R}^2} f_u(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2)^{\mathbf{b}}\pi_u^N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}} \\ &= \frac{\int_{\Omega_p}\int_{\mathbb{R}^2}\frac{f_u\left(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)\pi_u^N\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)}{\int_{\Omega_u}\int_{\mathbb{R}^2} f_u\left(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)\pi_u^N\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}{\int_{\Omega_p^a}\int_{\mathbb{R}^2}\frac{f_u(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2)^{\mathbf{b}}\pi_u^N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)}{\int_{\Omega_u}\int_{\mathbb{R}^2} f_u(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2)^{\mathbf{b}}\pi_u^N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}d\boldsymbol{\mu}d\boldsymbol{\sigma}^2} = \frac{\int_{\Omega_p}\int_{\mathbb{R}^2}\pi_u^F\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2|\mathbf{x}\right)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}{\int_{\Omega_p^a}\int_{\mathbb{R}^2}\pi_u^F\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2} \\ &= \frac{\int_{\Omega_p}\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}\right)d\boldsymbol{\sigma}^2}{\int_{\Omega_p^a}\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)d\boldsymbol{\sigma}^2} = \frac{P^F\left(\boldsymbol{\sigma}^2\in\Omega_p|\mathbf{x}\right)}{P^F\left(\boldsymbol{\sigma}^2\in\Omega_p^a|\mathbf{x}^{\mathbf{b}}\right)}\end{aligned} \tag{2.41}$$

Chapter 3

Bayesian Evaluation of Constrained Hypotheses on Variances of Multiple Independent Groups

Abstract

Research has shown that independent groups often differ not only in their means, but also in their variances. Comparing and testing variances is therefore of crucial importance to understand the effect of a grouping variable on an outcome variable. Researchers may have specific expectations concerning the relations between the variances of multiple groups. Such expectations can be translated into hypotheses with inequality and/or equality constraints on the group variances. Currently, however, no methods are available for testing (in)equality constrained hypotheses on variances. This chapter proposes a novel Bayesian approach to this challenging testing problem. Our approach has the following useful properties: First, it can be used to simultaneously test multiple (non)nested hypotheses with equality as well as inequality constraints on the variances. Second, our approach is fully automatic in the sense that no subjective prior specification is needed. Only the hypotheses need to be provided. Third, a user-friendly software application is included that can be used to perform this Bayesian test in an easy manner.

3.1 Introduction

Data analysis in psychological research commonly focuses on measures of central tendency such as means and regression coefficients. Measures of dispersion like variances receive relatively little attention. By disregarding the dispersion, however, researchers

This chapter is published as Böing-Messing, F., van Assen, M. A. L. M., Hofman, A. D., Hoijtink, H., & Mulder, J. (2017). Bayesian evaluation of constrained hypotheses on variances of multiple independent groups. *Psychological Methods*, *22*(2), 262–287. http://dx.doi.org/10.1037/met0000116.

run the risk of overlooking vital information in the data. Carroll (2003) distinguished two situations in which it is crucial to carefully consider the structure of variances. The first is the situation in which the variability systematically depends on known factors. An example is heteroscedasticity in ANOVA and regression. For instance, it has been pointed out that in experimental studies treatments often not only affect group means, but also group variances (e.g., Bryk & Raudenbush, 1988; Grissom, 2000; Ruscio & Roche, 2012). However, heterogeneity of variances is common in existing groups as well (e.g., Grissom, 2000; Ruscio & Roche, 2012). For example, males have been found to be more variable than females on a variety of measures (e.g., Lehre, Lehre, Laake, & Danbolt, 2009). Furthermore, it is frequently observed that the variability changes systematically with time (e.g., Aunola, Leskinen, Lerkkanen, & Nurmi, 2004; Hultsch, MacDonald, & Dixon, 2002). For example, a method that allows for the variability to systematically depend on known factors is the beta regression approach of Smithson and Verkuilen (2006). The authors model the mean as well as the variance as a function of (possibly different) predictors, thus treating the variance as a parameter of interest rather than as a nuisance parameter. The second situation in which variances play a crucial role is in multilevel modeling. Here researchers need to carefully model the variability at multiple levels, which results in multiple variance components. For example, Verhagen and Fox (2013) proposed a test on variance components in multilevel IRT models to check for measurement invariance in cross-national surveys. Furthermore, Kim and Seltzer (2011) examined heterogeneity in residual variance in multilevel models applied to (quasi-)experimental data in order to detect differential response to treatments. In the present chapter the focus is on heterogeneity of variances in one-way ANOVA designs with independent groups in the first situation.

There are often reasons to expect a certain structure of the variances of multiple independent groups. Typically one expects that certain groups are more heterogeneous than others, less heterogeneous, or equally heterogeneous. Such expectations can be translated into equality and inequality constrained hypotheses on the group variances. For example, in experimental studies one would expect treatment groups to have larger variances than control groups because participants respond differently to treatments (e.g., Bryk & Raudenbush, 1988; Grissom, 2000). Suppose we compare a control group with two treatment groups receiving a mild and an intense treatment, respectively. A conceivable hypothesis in this case would be $H_1\colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$, where σ_1^2 is the variance of the control group and σ_2^2 and σ_3^2 are the variances of the groups receiving the mild and the intense treatment, respectively. Note that H_1 states that the intense treatment produces larger variance than the mild treatment. To see whether there is evidence in favor of H_1 we test it against one or more competing hypotheses. Potential competitors are the null hypothesis $H_0\colon \sigma_1^2 = \sigma_2^2 = \sigma_3^2$ stating equality of variances and the complement of H_1 given by $H_2\colon$ not $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$. The complement H_2 entails all possible hypotheses except H_1. Hence, testing an order constrained hypothesis like H_1 against its complement tells us whether there is evidence in favor of our expected order or whether another hypothesis is more likely. Note that the interest is solely on the group variances, whereas the group means are treated as nuisance parameters.

Theoretical considerations often suggest (in)equality constrained hypotheses on the variances of existing groups as well. For example, Aunola et al. (2004) hypothe-

sized that the variance of mathematical abilities either increases or decreases across grades. For $J \geqslant 2$ grades this can be expressed in the two competing hypotheses $H_1\colon \sigma_1^2 < \cdots < \sigma_J^2$ and $H_2\colon \sigma_J^2 < \cdots < \sigma_1^2$, where σ_j^2 denotes the variance in grade j. The idea behind an increase (H_1) is that students who start out with high mathematical potential develop their mathematical abilities faster than students with low potential, which increases interindividual differences. A decrease in the variability of mathematical abilities (H_2) might occur because systematic instruction at school helps students with low mathematical potential catch up, so that interindividual differences decrease. Another potential competing hypothesis would be the null hypothesis $H_0\colon \sigma_1^2 = \cdots = \sigma_J^2$. Note that H_1 and H_2 are in agreement with models of development over time. For example, in the random slope model variances may increase over time, decrease over time, or first decrease and then increase over time (Snijders & Bosker, 2012). Constrained hypotheses on the variances of existing groups are conceivable in a variety of psychological research areas. For example, research on gender differences often finds males to be more variable in their intellectual abilities and personality than females (e.g., Borkenau et al., 2013; Feingold, 1992). Gerontological studies have found that the variability of reaction times increases with age (e.g., Hultsch et al., 2002). Research on psychological disorders has shown that ADHD patients tend to be more variable in their attentional performances than groups of people who do not suffer from ADHD (e.g., Silverstein, Como, Palumbo, West, & Osborn, 1995). Furthermore, research on person-in-context behavior suggests that the variability of people's behavior may differ across situations. For example, Van Mechelen (2009) argued that in an aggression context the variability may depend on the amount of social control in a situation, where high social control results in homogeneous behavior and thus low variability.

The standard approach to testing variances is null hypothesis significance testing (NHST). Classical NHST procedures like the likelihood ratio test or Levene's test (Levene, 1960) test the null hypothesis stating that all J variances are equal, $H_0\colon \sigma_1^2 = \cdots = \sigma_J^2$, against the alternative hypothesis stating that the variances are not all equal, $H_a\colon$ not $\sigma_1^2 = \cdots = \sigma_J^2$. In testing the order constrained hypothesis $H_1\colon \sigma_1^2 < \cdots < \sigma_J^2$ using classical NHST procedures one would proceed as follows: First we test the null against the alternative hypothesis. If we are able to reject the null hypothesis, we check whether the sample variances follow the order stated in the order constrained hypothesis. For more than two groups this is done by pairwise comparisons. This approach entails two problems: First, it suffers from Type I error inflation if we do not adjust the significance level for multiple testing. If we do adjust the significance level, then the procedure suffers from low power (e.g., Cohen, 1992). Second, it is possible that the pairwise comparisons produce contradictory results (e.g., $H_0\colon \sigma_1^2 = \sigma_2^2$ and $H_0\colon \sigma_1^2 = \sigma_3^2$ are not rejected, but $H_0\colon \sigma_2^2 = \sigma_3^2$ is).

Motivated by these disadvantages, Gastwirth, Gel, and Miao (2009) proposed an NHST procedure for testing the null hypothesis against an order constrained hypothesis. The advantage of this test is that it has higher power to detect an order effect. However, the method does not allow testing the null against an alternative hypothesis with a combination of equality and inequality constraints on the variances. This is a serious limitation given the large number of distinct hypotheses we can formulate. Using different combinations of equality and inequality constraints, we can specify dozens of distinct hypotheses on three variances. For more than three groups there

are well over 100 distinct hypotheses. Furthermore, the test by Gastwirth et al. does not solve the problems inherent in all NHST procedures: First, NHST procedures are not able to quantify evidence in favor of a hypothesis, no matter whether it is a null, an order constrained, or an unconstrained hypothesis (e.g., Wagenmakers, 2007). Second, it often happens that researchers have multiple competing hypotheses they would like to compare. NHST procedures do not allow testing these hypotheses against one another to determine which is most supported by the data. All one can do is test each hypothesis against the null, which does not answer the research question which hypothesis receives strongest support.

Given the problems with NHST procedures, it seems natural to use an information criterion like the Akaike information criterion (AIC; Akaike, 1973) or the Bayesian information criterion (BIC; Schwarz, 1978) to compare the hypotheses. However, these criteria cannot be used to test inequality constrained hypotheses. Both the AIC and the BIC involve a penalty term that measures the complexity of a hypothesis by the number of parameters. However, under inequality constrained hypotheses the number of parameters is not a suitable measure of the complexity because each inequality constraint effectively reduces the complexity. For example, the order constrained hypothesis $H_1\colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ is less complex than the unconstrained hypothesis $H_u\colon \sigma_1^2, \sigma_2^2, \sigma_3^2$ because under H_1 the variances can take on fewer values (e.g., σ_1^2 cannot be greater than σ_2^2). As a solution to this problem, Anraku (1999) proposed the order-restricted information criterion (ORIC). However, the ORIC is designed for testing order constrained hypotheses on means. At this point it is unclear whether this methodology can be generalized to the case of testing equality and inequality constrained hypotheses on variances. Note that the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002) and the Watanabe-Akaike information criterion (WAIC; Watanabe, 2010) do not provide a solution to this problem because they do not properly take the parsimony introduced by inequality constraints into account (Mulder et al., 2009; Gelman, Hwang, & Vehtari, 2014). Under certain conditions the DIC and the WAIC are asymptotically equal to leave-one-out cross-validation (Gelman et al., 2014), which implies that the latter is not suitable for testing inequality constrained hypotheses on variances either.

In this chapter we adopt a Bayesian approach to testing equality and inequality constrained hypotheses on variances using Bayes factors (Jeffreys, 1961; Kass & Raftery, 1995). The Bayes factor is a Bayesian hypothesis testing and model selection criterion. It provides a solution to the aforementioned problems inherent in NHST procedures and existing information criteria. In particular, the Bayes factor quantifies the evidence in favor of a hypothesis. This holds for all types of hypotheses: The Bayes factor allows quantification of evidence in favor of a null hypothesis, order constrained hypotheses, and hypotheses with a combination of equality and inequality constraints. Furthermore, using the Bayes factor it is straightforward to simultaneously test multiple hypotheses against one another. In this case the Bayes factor tells us which hypothesis is most supported by the data. Bayes factors have a number of additional desirable properties: First, contrary to NHST procedures, Bayes factors do not require the hypotheses under consideration to be nested (e.g., Berger & Mortera, 1999). Bayes factors are therefore able to directly test, for example, $H_1\colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ against $H_2\colon \sigma_3^2 < \sigma_2^2 < \sigma_1^2$. Second, Bayes factors automatically function as Occam's razor. This means that if two hypotheses describe the data equally well, then the

Bayes factor automatically chooses the more parsimonious hypothesis. This is a useful property of Bayes factors because it is frequently observed that parsimonious hypotheses that describe the data well are more likely to be correct than complex ones. Third, Bayes factors are consistent. This means that the Bayes factor always chooses the true hypothesis if we have enough data.

Bayes factors have been developed for various testing problems frequently encountered in the psychological sciences. For instance, Rouder, Speckman, Sun, Morey, and Iverson (2009) proposed a Bayesian t-test. Klugkist, Laudy, and Hoijtink (2005) discussed a Bayes factor for testing hypotheses on mean parameters in analysis of variance designs. Mulder, Hoijtink, and Klugkist (2010) presented methods for Bayesian testing of means and regression coefficients in the multivariate normal linear model. Gu, Mulder, Deković, and Hoijtink (2014) proposed an approximate Bayes factor for evaluating hypotheses with inequality constraints on means and regression parameters. In the present chapter we propose a novel Bayes factor for testing equality and inequality constrained hypotheses on variances of multiple independent groups. Our methodology builds upon the fractional Bayes factor of O'Hagan (1995) in combination with the prior adjustment of Mulder (2014b) and Böing-Messing and Mulder (2016).

The remainder of this chapter is structured as follows. First, we discuss the statistical model and options for formulating hypotheses on the group variances. We then give a brief introduction to the Math Garden (Klinkenberg, Straatemeier, & van der Maas, 2011; Straatemeier, 2014), which we use to illustrate the importance of testing (in)equality constrained hypotheses on variances. Next, we discuss Bayes factors for testing hypotheses on variances. We first apply the fractional Bayes factor (O'Hagan, 1995) to the testing problem and show that it may not function as Occam's razor when testing inequality constrained hypotheses. As a novel solution to this problem we propose an adjusted fractional Bayes factor. The performance of the new method is illustrated in a simulation study. Following this, we continue the illustrative example by applying the adjusted fractional Bayes factor to data from the Math Garden. We then present a user-friendly software application for computing the adjusted fractional Bayes factor. We conclude the chapter with a discussion of our approach.

3.2 Model and Hypotheses

We consider the one-way ANOVA design with $J \geqslant 2$ independent groups of size n_j, $j = 1, \ldots, J$. Each observation in group j is assumed to be independent and normally distributed with mean μ_j and variance σ_j^2. The unconstrained likelihood with no constraints on the group means and variances is given by

$$f_u\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) = \prod_{j=1}^{J} \prod_{i=1}^{n_j} N\left(x_{ij}|\mu_j, \sigma_j^2\right), \tag{3.1}$$

where $\mathbf{x}$ are the data, x_{ij} is the ith observation from the jth group, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_J)'$ is the vector of group means, and $\boldsymbol{\sigma}^2 = \left(\sigma_1^2, \ldots, \sigma_J^2\right)'$ is the vector of group variances.

Hypotheses on the variances can be formulated using two basic types of constraints: equality constraints and inequality constraints. With equality constraints

we can specify equalities of two or more variances, for example $H\colon \sigma_1^2 = \sigma_2^2 = \sigma_3^2$. Inequality constraints are used to formulate expectations regarding differences in magnitude between variances, for example $H\colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$. If we do not expect certain relations between variances, then we simply do not impose constraints on them. We shall use the comma symbol (,) to indicate that there are no constraints between variances, for example $H_u\colon \sigma_1^2, \sigma_2^2, \sigma_3^2$. We refer to the hypothesis with no constraints on the variances as the unconstrained hypothesis. In formulating hypotheses we may combine equality constraints, inequality constraints, and no constraints between variances, for example $H\colon \sigma_1^2 = \sigma_2^2 < \sigma_3^2, \sigma_4^2$. Another hypothesis that is often of interest is the complement of an order constrained hypothesis. For example, the complement of the order constrained hypothesis $H_1\colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ is given by $H_2\colon$ not $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$, for which we also write $H_2\colon$ not H_1 in short. The complement entails all possible hypotheses except the order constrained hypothesis. We may also test the complement of multiple orders. For example, Aunola et al. (2004) expected the variance of mathematical abilities to either increase or decrease across grades. This corresponds to the two order constrained hypotheses $H_1\colon \sigma_1^2 < \cdots < \sigma_J^2$ and $H_2\colon \sigma_J^2 < \cdots < \sigma_1^2$, for which the complement is given by $H_3\colon$ not (H_1 or H_2). Note that one may also perform the classical test of the null hypothesis $H_0\colon \sigma_1^2 = \cdots = \sigma_J^2$ against the unconstrained alternative hypothesis $H_u\colon \sigma_1^2, \ldots, \sigma_J^2$ if the interest is on whether the group variances are equal or not. The likelihood under a constrained hypothesis H_t is a truncation of the unconstrained likelihood in Equation (3.1) in the parameter space that is admissible under H_t, which we denote by Ω_t:

$$f_t\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) = f_u\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) I_{\Omega_t}\left(\boldsymbol{\sigma}^2\right), \tag{3.2}$$

where $I_{\Omega_t}\left(\boldsymbol{\sigma}^2\right)$ is an indicator function that equals 1 if the variances $\boldsymbol{\sigma}^2$ are in the admissible parameter space Ω_t and 0 if the variances are outside the admissible parameter space.

3.3 Illustrative Example: The Math Garden

The Math Garden (Klinkenberg et al., 2011; Straatemeier, 2014) is an online adaptive learning environment for basic mathematics. It is currently used by more than 300,000 children in primary education, involving more than 4,000 schools. Next to providing children and teachers with an online learning tool, the system opens up a valuable data base for researchers. In this chapter we present an analysis of children's abilities in four different games, each covering one of the basic mathematical operations addition, subtraction, multiplication, and division.

As mentioned in the introduction, Aunola et al. (2004) hypothesized that the variance of mathematical abilities either increases or decreases across grades. This suggests testing the following two research hypotheses in the Math Garden:

$$\begin{aligned} H_1&\colon \sigma_1^2 < \cdots < \sigma_J^2, \\ H_2&\colon \sigma_J^2 < \cdots < \sigma_1^2, \end{aligned} \tag{3.3}$$

where σ_j^2 is the variance of mathematical abilities in grade j and J is the number of grades to be compared. Thus, H_1 states an increase in variance, whereas H_2

states a decrease. We shall test these two research hypotheses against two competing hypotheses:

$$\begin{aligned} &H_0\colon \sigma_1^2 = \cdots = \sigma_J^2, \\ &H_3\colon \text{not } (H_0 \text{ or } H_1 \text{ or } H_2). \end{aligned} \tag{3.4}$$

Here H_0 is the classical null hypothesis stating equality of variances. Hypothesis H_3 is the complement of H_0, H_1, and H_2. We include it to cover all possible hypotheses in case neither the research hypotheses nor the null hypothesis is supported by the data. In the Math Garden a player's ability is estimated separately for each of the four games addition, subtraction, multiplication, and division. That is, each player has a separate ability estimate for each game they play. We will therefore test the hypotheses in Equations (3.3) and (3.4) for each game separately.

3.4 Bayes Factors for Testing Constrained Hypotheses on Variances

The Bayes factor is a Bayesian testing criterion that can be used to quantify the relative evidence in the data between two hypotheses. The main ingredient of the Bayes factor is the marginal likelihood of the data under each hypothesis. The marginal likelihood of the data $\mathbf{x}$ under the constrained hypothesis H_t, denoted by m_t, is defined by the integral over the product of the likelihood, denoted by f_t, and the prior, denoted by π_t, over the admissible parameter space under H_t. The marginal likelihood can be expressed as

$$m_t(\mathbf{x}) = \int_{\Omega_t} \int_{\mathbb{R}^J} f_t\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) \pi_t\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2, \tag{3.5}$$

where the likelihood f_t under the constrained hypothesis H_t was given in Equation (3.2), the prior distribution π_t contains the information about the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$ before observing the data, which will be discussed below, and Ω_t denotes the constrained parameter space of the variances under H_t. For example, for $H_1\colon \sigma_1^2 < \cdots < \sigma_J^2$ the constrained space Ω_1 corresponds to the subspace of the variances that is in agreement with the ordering $\sigma_1^2 < \cdots < \sigma_J^2$. The marginal likelihood quantifies how well the model and the prior under H_t were able to predict the observed data (Jeffreys, 1961; Kass & Raftery, 1995).

In order to determine the evidence in the data in favor of a hypothesis, say H_1, relative to another hypothesis, say H_2, the ratio of the marginal likelihoods needs to be computed via

$$B_{12} = \frac{m_1(\mathbf{x})}{m_2(\mathbf{x})}, \tag{3.6}$$

which is known as the Bayes factor of hypothesis H_1 against hypothesis H_2. If the Bayes factor B_{12} is larger (smaller) than 1, this indicates that the evidence in the data in favor of H_1 (H_2) is stronger than the evidence in favor of H_2 (H_1). For example, a Bayes factor of $B_{12} = 10$ implies that the evidence in the data in favor of H_1 is ten times as strong as the evidence in favor of H_2. Kass and Raftery (1995) provided interpretation guidelines for the Bayes factor as stated in Table 3.1. We would like

Table 3.1: Interpretation guidelines for the Bayes factor B_{12} testing hypothesis H_1 against hypothesis H_2 (from Kass & Raftery, 1995).

B_{12}	Evidence in favor of H_1
1 to 3	Not worth more than a bare mention
3 to 20	Positive
20 to 150	Strong
> 150	Very strong

to emphasize, however, that these guidelines should not be used as strict rules when interpreting Bayes factors. A researcher should decide for himself or herself whether a Bayes factor of, say, $B_{12} = 120$ is enough to completely rule out hypothesis H_2 in comparison to hypothesis H_1.

Prior specification is an important step when computing the marginal likelihood. First, it is important to note that priors should not be specified in an ad hoc manner because the Bayes factor strongly depends on the exact choice of the prior. For instance, the Bayes factor for a null hypothesis against an unconstrained alternative hypothesis can be made arbitrarily large when specifying the prior under the unconstrained alternative extremely vague. This is known as Bartlett's phenomenon (e.g., Bartlett, 1957; Jeffreys, 1961; Liang, Paulo, Molina, Clyde, & Berger, 2008; Lindley, 1957). Alternatively, one might consider using noninformative improper priors, which are commonly used in objective Bayesian estimation (Berger, 2006). When using Bayes factors, however, it is not possible to work with noninformative improper priors since these contain undefined normalizing constants which do not cancel out when computing the marginal likelihoods and Bayes factors according to Equations (3.5) and (3.6).

Thus, in order to quantify the relative evidence in the data between constrained hypotheses on variances using the Bayes factor one needs to carefully formulate proper priors for the unknown parameters under all hypotheses under consideration. For instance, in the Math Garden example a proper prior needs to be specified for the group variances under H_1 satisfying the increasing order, the group variances under H_2 satisfying the decreasing order, the common group variance under H_0, and the group variances under the complement hypothesis H_3. Because often precise prior information about the degree of heterogeneity across populations is not available, specification of proper priors is a difficult task for a researcher. This holds especially when testing hypotheses with constraints on variances.

To avoid this limitation statisticians have developed automatic (or default) marginal likelihoods and Bayes factors that enable researchers to automatically quantify the relative evidence in the data between the hypotheses. These default Bayes factors can be computed in an automatic fashion without needing to specify proper priors for the model parameters based on one's subjective prior beliefs. Well-known examples are the fractional Bayes factor (O'Hagan, 1995), the intrinsic Bayes factor (Berger & Pericchi, 1996), and the Bayes factor based on expected-posterior priors (Mulder et al., 2009; Pérez & Berger, 2002). Here we shall focus on the fractional Bayes factor because it is computationally efficient and has desirable theoretical prop-

erties (O'Hagan, 1995, 1997).

3.4.1 Fractional Bayes Factors

The fractional Bayes factor (FBF) was proposed by O'Hagan (1995) to circumvent the need to specify a proper prior based on external prior information. In the FBF, the marginal likelihood is defined as

$$m_t^F(\mathbf{x}, b) = \frac{\int_{\Omega_t} \int_{\mathbb{R}^J} f_t\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) \pi_t^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}{\int_{\Omega_t} \int_{\mathbb{R}^J} f_t\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^b \pi_t^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}, \tag{3.7}$$

where π_t^N denotes a noninformative improper prior and the fraction b can take on values between 0 and 1. Thus, the marginal likelihood in the FBF corresponds to the standard marginal likelihood based on a noninformative improper prior divided by the standard marginal likelihood where the likelihood is raised to the power of the fraction b. O'Hagan (1995) motivates this form of the marginal likelihood in the context of partial Bayes factors. In particular, he argues that the fraction of the likelihood (i.e., the likelihood to the power of b) contains a part of the information in the full likelihood in the sense that the fraction of the likelihood is approximately equal to the likelihood based on a training sample if we set $b = m/n$, where both the sample size n and the training sample size m are large. As will be elaborated below, the fraction b controls the amount of information in the implicit automatic proper prior.

The noninformative improper prior we use in Equation (3.7) is the standard independence Jeffreys prior. For a constrained hypothesis, this noninformative improper prior is proportional to the product of the reciprocals of the unique variances truncated in the inequality constrained parameter space (if there are inequality constraints present). For example, under $H_0\colon \sigma_1^2 = \cdots = \sigma_J^2$ with one unique variance, say, σ^2, and $H_1\colon \sigma_1^2 < \cdots < \sigma_J^2$ with J unique variances that are inequality constrained, the noninformative improper priors are given by

$$\begin{aligned} \pi_0^N\left(\boldsymbol{\mu}, \sigma^2\right) &= C_0 \times \sigma^{-2} \quad \text{and} \\ \pi_1^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) &= C_1 \times \sigma_1^{-2} \times \cdots \times \sigma_J^{-2} \times I\left(\sigma_1^2 < \cdots < \sigma_J^2\right), \end{aligned} \tag{3.8}$$

respectively, where $I\left(\sigma_1^2 < \cdots < \sigma_J^2\right)$ is an indicator function that equals 1 if $\sigma_1^2 < \cdots < \sigma_J^2$ and 0 otherwise, and C_0 and C_1 denote the respective undefined normalizing constants. Because the noninformative improper prior appears in the numerator as well as in the denominator in the marginal likelihood in Equation (3.7), the undefined constants in the improper prior cancel out in the FBF approach. Note that the noninformative priors imply flat priors for the group means.

The fraction b controls how much of the information in the data is used to specify an automatic proper prior. This can be made explicit by rewriting the marginal

likelihood in Equation (3.7) following Gilks (1995):

$$\begin{aligned} m_t^F(\mathbf{x}, b) &= \int_{\Omega_t} \int_{\mathbb{R}^J} f_t\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{1-b} \frac{f_t\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^b \pi_t^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)}{\int_{\Omega_t} \int_{\mathbb{R}^J} f_t\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^b \pi_t^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2} d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2 \\ &= \int_{\Omega_t} \int_{\mathbb{R}^J} f_t\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{1-b} \pi_t\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2|\mathbf{x}^b\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2, \end{aligned} \tag{3.9}$$

where

$$\pi_t\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2|\mathbf{x}^b\right) = \frac{f_t\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^b \pi_t^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)}{\int_{\Omega_t} \int_{\mathbb{R}^J} f_t\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^b \pi_t^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2} \tag{3.10}$$

is the automatic proper prior that is obtained by updating the noninformative improper prior with a fraction b of the likelihood. Note that the symbol $\mathbf{x}^b$ is used to illustrate that this prior contains a fraction b of the information in the complete data $\mathbf{x}$. As can be seen from Equation (3.9), in computing the marginal likelihood the fraction b of the likelihood is used to obtain a proper automatic prior and the remaining fraction $1 - b$ is used for hypothesis testing. It is generally recommendable to choose the fraction b based on the minimal number of observations that is needed to obtain a proper automatic prior when updating the improper Jeffreys prior (e.g., Berger & Mortera, 1999; O'Hagan, 1995). In our testing problem with $2J$ unknown parameters (i.e., J unknown means and J unknown variances), we need at least $2J$ observations to obtain a proper prior when updating the improper Jeffreys prior. This implies setting $b = 2J/N$, where $N = \sum_{j=1}^{J} n_j$ is the total sample size. This choice ensures that the remaining fraction $1 - b$ that is used for hypothesis testing is maximal. As was shown by O'Hagan (1995), the FBF is consistent under very general settings, which implies that as the sample size grows to infinity, the evidence in favor of the true hypothesis goes to infinity. If we use a minimal fraction the evidence in favor of the true hypothesis goes fastest to infinity, which makes the minimal fraction the optimal choice.

3.4.2 Fractional Bayes Factors for an Inequality Constrained Test

Next we apply the FBF to test the inequality constrained hypothesis $H_1\colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ against the unconstrained hypothesis $H_u\colon \sigma_1^2, \sigma_2^2, \sigma_3^2$. The inequality constrained subspace under H_1 can be written as $\Omega_1 = \left\{\boldsymbol{\sigma}^2 | \sigma_1^2 < \sigma_2^2 < \sigma_3^2\right\}$. As shown in Appendix 3.A, the FBF for H_1 against H_u can be written as the posterior probability that the constraints of H_1 hold divided by the automatic prior probability that the constraints of H_1 hold:

$$B_{1u}^F = \frac{m_1^F(\mathbf{x}, b)}{m_u^F(\mathbf{x}, b)} = \frac{P\left(\boldsymbol{\sigma}^2 \in \Omega_1|\mathbf{x}\right)}{P\left(\boldsymbol{\sigma}^2 \in \Omega_1|\mathbf{x}^b\right)} = \frac{P\left(\sigma_1^2 < \sigma_2^2 < \sigma_3^2|\mathbf{x}\right)}{P\left(\sigma_1^2 < \sigma_2^2 < \sigma_3^2|\mathbf{x}^b\right)}. \tag{3.11}$$

The unconstrained marginal automatic prior for the variances, which is needed to compute the probability in the denominator in Equation (3.11), can be obtained by

integrating the group means out of the joint automatic prior:

$$\pi_u\left(\boldsymbol{\sigma}^2|\mathbf{x}^b\right) = \int_{\mathbb{R}^3} \pi_u\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2|\mathbf{x}^b\right) d\boldsymbol{\mu} = \prod_{j=1}^{3} \text{Inv-}\chi^2\left(\sigma_j^2 \,\middle|\, bn_j - 1, \frac{b(n_j-1)s_j^2}{bn_j-1}\right), \quad (3.12)$$

which is a product of scaled inverse-χ^2 distributions with degrees of freedom of $bn_j - 1$ and scale hyperparameters of $\frac{b(n_j-1)s_j^2}{bn_j-1}$, where $s_j^2 = \frac{1}{n_j-1}\sum_{i=1}^{n_j}\left(x_{ij} - \bar{x}_j\right)^2$ is the sample variance of group j. In this setting the minimal fraction is given by $b = 6/N$, where $N = n_1+n_2+n_3$. The unconstrained marginal posterior can simply be obtained by plugging $b = 1$ into Equation (3.12), which yields

$$\pi_u\left(\boldsymbol{\sigma}^2|\mathbf{x}\right) = \prod_{j=1}^{3} \text{Inv-}\chi^2\left(\sigma_j^2|n_j - 1, s_j^2\right). \quad (3.13)$$

The distributions above can be used to obtain a large sample of, say, $S = 100{,}000$ draws from the unconstrained posterior and unconstrained automatic prior (see Gelman et al., 2004, for information on how to sample from the scaled inverse-χ^2 distribution). Subsequently, by taking the proportion of unconstrained draws that satisfy the constraints of H_1, the fractional Bayes factor in Equation (3.11) can be computed as

$$B_{1u}^F \approx \frac{S^{-1}\sum_{s=1}^{S} I\left(\sigma_{1,post}^{2(s)} < \sigma_{2,post}^{2(s)} < \sigma_{3,post}^{2(s)}\right)}{S^{-1}\sum_{s=1}^{S} I\left(\sigma_{1,prior}^{2(s)} < \sigma_{2,prior}^{2(s)} < \sigma_{3,prior}^{2(s)}\right)}, \quad (3.14)$$

where $\boldsymbol{\sigma}_{post}^{2(s)} = \left(\sigma_{1,post}^{2(s)}, \sigma_{2,post}^{2(s)}, \sigma_{3,post}^{2(s)}\right)'$ and $\boldsymbol{\sigma}_{prior}^{2(s)} = \left(\sigma_{1,prior}^{2(s)}, \sigma_{2,prior}^{2(s)}, \sigma_{3,prior}^{2(s)}\right)'$ are the sth draw from the unconstrained posterior and automatic prior, respectively, for $s = 1, \ldots, S$.

It is important to note that the use of a common fraction b for all groups may be problematic in the case of unbalanced data with unequal group sizes. For example, when $n_1 = 10$, $n_2 = 20$, and $n_3 = 30$, it holds that $b = 6/60 = 0.1$. This results in prior degrees of freedom of 0, 1, and 2, for σ_1^2, σ_2^2, and σ_3^2, respectively. However, the degrees of freedom must be larger than 0. This shows that the standard FBF approach is not generally applicable in the case of unequal group sizes. We come back to this issue in the next section.

Another important consequence of using a fraction of the data for constructing the automatic prior in Equation (3.12) is that the scale hyperparameter of each variance σ_j^2 depends on the corresponding sample variance s_j^2. This implies that the automatic prior is concentrated around the observed effect, which has undesirable consequences when testing inequality constrained hypotheses on variances. We illustrate this with an example. For the moment, let us consider a balanced data set with equal group sizes of $n_j = n = 20$, for $j = 1, 2$, and 3, and let the sample variances satisfy $s_1^2 = 1$, $s_2^2 = s$, and $s_3^2 = s^2$. Thus, if $s > 1$, then there is evidence in favor of H_1 because the sample variances are in agreement with the inequality constraints under H_1. Similarly, if $s < 1$, then there is evidence against H_1 because the sample variances are not in agreement with the inequality constraints. Note that the degrees of freedom in the automatic prior equal $6/60 \times 20 - 1 = 1$, which implies a distribution with minimal information.

Figure 3.1 shows the FBF for H_1 against H_u (solid line) when letting s^2 increase from $\exp(-10) \approx 0.00$ to $\exp(10) \approx 22{,}000$. As s^2 becomes large (which implies clear evidence in favor of H_1), the FBF goes to 1. This can be explained by the fact that as s^2 increases, the unconstrained posterior in Equation (3.13) as well as the unconstrained automatic prior in Equation (3.12) become completely located in the constrained space of H_1. For example, in Figure 3.2a it can be seen that a large portion of an isodensity surface of the automatic prior for $s^2 = 9$ and $n_j = 20$ is located in the inequality constrained space $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$ (marked with thick lines). The automatic prior probability that the inequality constraints hold is equal to 0.38 in this case. As s^2 increases, both the posterior and the prior probability that the inequality constraints hold go to 1 because the posterior and the automatic prior become completely located in the inequality constrained space. Therefore the ratio of the two probabilities in Equation (3.11) also goes to 1. Thus, in the FBF approach the parsimonious order constrained hypothesis that is strongly supported by the data does not receive stronger support than the more complex unconstrained hypothesis. This implies that the FBF does not function as Occam's razor in this situation. This undesirable property is a direct consequence of the fact that the automatic prior for the group variances is concentrated around the sample variances. For this reason we propose an adjustment of the FBF that corrects for this undesirable behavior when testing inequality constrained hypotheses on variances.

3.4.3 Adjusted Fractional Bayes Factors

In this section we present two novel extensions of the FBF approach for testing hypotheses with equality and inequality constraints on variances. The resulting criterion will be referred to as the adjusted fractional Bayes factor (aFBF).

In the aFBF the marginal likelihood is defined as

$$m_t^{aF}(\mathbf{x}, \mathbf{b}) = \frac{\int_{\Omega_t} \int_{\mathbb{R}^J} f_u\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) \pi_u^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}{\int_{\Omega_t^a} \int_{\mathbb{R}^J} f_u\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}} \pi_u^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}. \tag{3.15}$$

This expression has three important differences in comparison to the marginal likelihood in the FBF approach given in Equation (3.7). First, in the denominator in Equation (3.15) we integrate over an adjusted parameter space, which is denoted by Ω_t^a. The adjusted parameter space contains the same constraints as the unadjusted space Ω_t, except that each variance σ_j^2 is multiplied by a tuning parameter a_j. These tuning parameters are chosen such the automatic prior probability that the inequality constraints hold is based on prior distributions for the variances with equal scale hyperparameters (unlike in the FBF, as was observed in Equation (3.12)). Details on the choice of the tuning parameters will be discussed below. This adjustment results in a criterion that always incorporates the parsimony of a hypothesis with inequality constraints on the variances (Böing-Messing & Mulder, 2016; Mulder, 2014b). Second, in the denominator in Equation (3.15) the fraction of the likelihood is based on group-specific fractions $\mathbf{b} = (b_1, \ldots, b_J)'$, where the fraction of the likelihood of group j depends on the group size according to $b_j = 2/n_j$, for $j = 1, \ldots, J$. This generalization ensures that the minimal amount of information based on two observations per group is used for automatic prior specification. This was suggested by Berger

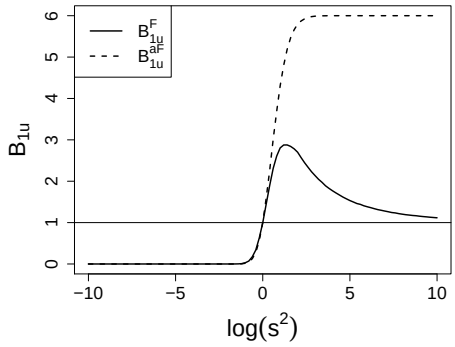
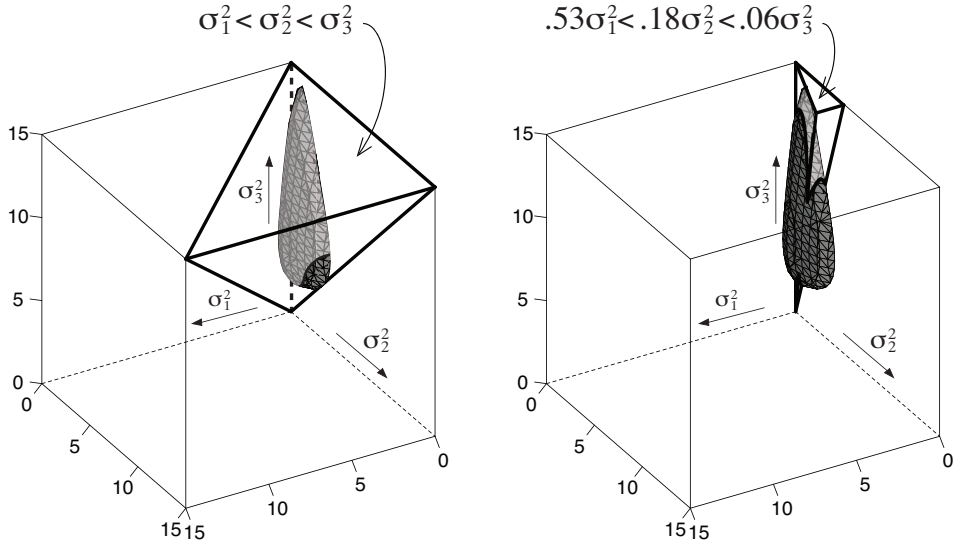


Figure 3.1: Fractional Bayes factor B^F_{1u} (solid line) and adjusted fractional Bayes factor B^{aF}_{1u} (dashed line) for testing $H_1\colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ against $H_u\colon \sigma_1^2, \sigma_2^2, \sigma_3^2$. The Bayes factors are plotted as a function of the sample variances $\left(s_1^2, s_2^2, s_3^2\right)' = \left(1, s, s^2\right)'$, where $s^2 \in [\exp(-10), \exp(10)]$, and for equal sample sizes of $n_1 = n_2 = n_3 = 20$.



(a) Unadjusted space

(b) Adjusted space

Figure 3.2: An isodensity surface of the automatic prior in Equation (3.12) for sample variances of $\left(s_1^2, s_2^2, s_3^2\right)' = (1, 3, 9)'$ and sample sizes of $n_1 = n_2 = n_3 = 20$. In Figure (a) the unadjusted parameter subspace satisfying $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$ is marked with thick lines. The automatic prior probability that the inequality constraints hold equals $P\left(\sigma_1^2 < \sigma_2^2 < \sigma_3^2 | \mathbf{x}^b\right) = 0.38$. Figure (b) shows the adjusted parameter subspace satisfying $0.53\sigma_1^2 < 0.18\sigma_2^2 < 0.06\sigma_3^2$ (marked with thick lines). The adjusted automatic prior probability that the inequality constraints hold equals $P\left(0.53\sigma_1^2 < 0.18\sigma_2^2 < 0.06\sigma_3^2 | \mathbf{x}^b\right) = 1/6$.

and Pericchi (2001) and De Santis and Spezzaferri (2001) for testing equality constraints on group means. Here we extend the idea to testing equality and inequality constrained hypotheses on variances. Finally, it is important to note that in the denominator in Equation (3.15) the likelihood and noninformative improper prior under the unconstrained hypothesis, f_u and π_u^N, are used instead of the likelihood and prior under the constrained hypothesis, f_t and π_t^N. This ensures that we integrate over the complete adjusted parameter space Ω_t^a in the denominator. For completeness, the unconstrained likelihood and prior are also used in the numerator of the marginal likelihood in the aFBF approach in Equation (3.15).

After some algebra (see Appendix 3.B for a proof) the marginal likelihood in the aFBF can be expressed as

$$m_t^{aF}(\mathbf{x}, \mathbf{b}) = \tilde{m}_t^{aF}(\mathbf{x}, \mathbf{b}) \frac{P\left(\boldsymbol{\sigma}^2 \in \Omega_t | \mathbf{x}\right)}{P\left(\boldsymbol{\sigma}^2 \in \Omega_t^a | \mathbf{x}^{\mathbf{b}}\right)}, \tag{3.16}$$

where

$$\begin{aligned}\tilde{m}_t^{aF}(\mathbf{x}, \mathbf{b}) = &\left(\prod_{k=1}^{K} \prod_{j=1}^{J_k} b_{k_j}^{\frac{1}{2}}\right) \pi^{-\frac{\sum_{k=1}^{K} \sum_{j=1}^{J_k} \left(1 - b_{k_j}\right) n_{k_j}}{2}} \\ &\prod_{k=1}^{K} \Gamma\left(\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}{2}\right) \Gamma\left(\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}{2}\right)^{-1} \\ &\left(\sum_{j=1}^{J_k} \left(n_{k_j} - 1\right) s_{k_j}^2\right)^{-\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}{2}} \left(\sum_{j=1}^{J_k} b_{k_j} \left(n_{k_j} - 1\right) s_{k_j}^2\right)^{\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}{2}}.\end{aligned} \tag{3.17}$$

In Equation (3.17) the expression $\Gamma(\cdot)$ is the gamma function, K denotes the number of unique variances, and J_k denotes the number of groups sharing the unique variance σ_k^2, for $k = 1, \ldots, K$. Furthermore, b_{k_j} and n_{k_j} are the fraction and the sample size of the jth group sharing the unique variance σ_k^2, for $j = 1, \ldots, J_k$. In Equation (3.16) the adjusted parameter space Ω_t^a is defined by

$$\Omega_t^a = \left\{ \boldsymbol{\sigma}^2 \middle| \left(a_1 \sigma_1^2, \ldots, a_K \sigma_K^2\right)' \in \Omega_t \right\}, \tag{3.18}$$

where the tuning parameters a_k are given by

$$a_k = \frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}{\sum_{j=1}^{J_k} b_{k_j} \left(n_{k_j} - 1\right) s_{k_j}^2}, \tag{3.19}$$

for $k = 1, \ldots, K$. Furthermore, the expressions $P\left(\boldsymbol{\sigma}^2 \in \Omega_t | \mathbf{x}\right)$ and $P\left(\boldsymbol{\sigma}^2 \in \Omega_t^a | \mathbf{x}^{\mathbf{b}}\right)$ are the posterior and the adjusted automatic prior probability that the inequality constraints on the variances hold, respectively. These can be computed by drawing a large sample of, say, $S = 100{,}000$ draws from the unconstrained posterior and automatic prior distribution of the variances given by

$$\pi_u\left(\boldsymbol{\sigma}^2 | \mathbf{x}\right) = \prod_{k=1}^{K} \text{Inv-}\chi^2\left(\sigma_k^2 \middle| \left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k, \frac{\sum_{j=1}^{J_k} \left(n_{k_j} - 1\right) s_{k_j}^2}{\left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}\right) \tag{3.20}$$

and

$$\pi_u\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) = \prod_{k=1}^{K} \text{Inv-}\chi^2\left(\sigma_k^2 \,\middle|\, \left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k, \frac{\sum_{j=1}^{J_k} b_{k_j}\left(n_{k_j}-1\right)s_{k_j}^2}{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}\right), \quad (3.21)$$

respectively. The posterior probability that the inequality constraints hold is then given by the proportion of posterior draws that satisfy the constraints, that is,

$$P\left(\boldsymbol{\sigma}^2 \in \Omega_t|\mathbf{x}\right) \approx \frac{1}{S}\sum_{s=1}^{S} I_{\Omega_t}\left(\boldsymbol{\sigma}_{post}^{2(s)}\right), \quad (3.22)$$

where $\boldsymbol{\sigma}_{post}^{2(s)}$ is the sth draw from the posterior in Equation (3.20), for $s = 1, \ldots, S$. Similarly, the adjusted prior probability that the inequality constraints hold is given by

$$P\left(\boldsymbol{\sigma}^2 \in \Omega_t^a|\mathbf{x}^{\mathbf{b}}\right) \approx \frac{1}{S}\sum_{s=1}^{S} I_{\Omega_t^a}\left(\boldsymbol{\sigma}_{prior}^{2(s)}\right), \quad (3.23)$$

where $\boldsymbol{\sigma}_{prior}^{2(s)}$ is the sth draw from the prior in Equation (3.21), for $s = 1, \ldots, S$.

Finally, it is important to note that the aFBF is scale invariant, that is, it does not depend on the scale of the outcome variable (a proof is given in Appendix 3.C). Note that scale invariance is of crucial importance because in comparing educational performances in different grades, for example, it should not matter whether students' performances are rated on a scale from 0 to 10 or from 0 to 100.

3.4.4 Adjusted Fractional Bayes Factors for an Inequality Constrained Test

Now we apply the aFBF to the test of $H_1\colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ against $H_u\colon \sigma_1^2, \sigma_2^2, \sigma_3^2$. As noted above, the adjusted parameter space contains the same constraints as the unadjusted space, except that the variances are multiplied by tuning parameters a_j which correct for the differences between the observed sample variances. Thus, the adjusted parameter space under H_1 is given by $\Omega_1^a = \left\{\boldsymbol{\sigma}^2 | a_1\sigma_1^2 < a_2\sigma_2^2 < a_3\sigma_3^2\right\}$, with $a_j = n_j/\left(2(n_j-1)s_j^2\right)$. Furthermore, the fractions are given by $\mathbf{b} = (b_1, b_2, b_3)'$, with $b_j = 2/n_j$, for $j = 1, 2$ and 3. The aFBF for H_1 against H_u can then be written as

$$B_{1u}^{aF} = \frac{P\left(\boldsymbol{\sigma}^2 \in \Omega_1|\mathbf{x}\right)}{P\left(\boldsymbol{\sigma}^2 \in \Omega_1^a|\mathbf{x}^{\mathbf{b}}\right)} = \frac{P\left(\sigma_1^2 < \sigma_2^2 < \sigma_3^2|\mathbf{x}\right)}{P\left(a_1\sigma_1^2 < a_2\sigma_2^2 < a_3\sigma_3^2|\mathbf{x}^{\mathbf{b}}\right)}. \quad (3.24)$$

Note that the posterior probability in the numerator is identical to that in the FBF in Equation (3.11). On the other hand, the automatic prior probability of the adjusted ordering in the denominator is computed using the automatic prior distribution

$$\begin{aligned}\pi_u\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) &= \prod_{j=1}^{3} \text{Inv-}\chi^2\left(\sigma_j^2 \,\middle|\, b_j n_j - 1, \frac{b_j(n_j-1)s_j^2}{b_j n_j - 1}\right) \\ &= \prod_{j=1}^{3} \text{Inv-}\chi^2\left(\sigma_j^2 \,\middle|\, 1, \frac{2(n_j-1)s_j^2}{n_j}\right).\end{aligned} \quad (3.25)$$

First note that the prior degrees of freedom are equal to 1, which implies minimal information for any group size n_j. Regarding the scale hyperparameter, standard mathematical statistics dictates that multiplying a random variable having a scaled inverse-χ^2 distribution by a constant, say a, results in a new random variable having a scaled inverse-χ^2 distribution where the original scale parameter is multiplied by a. For this reason, since $\sigma_j^2|\mathbf{x}^{\mathbf{b}} \sim \text{Inv-}\chi^2\left(1, 2(n_j-1)s_j^2/n_j\right)$, it automatically holds that

$$a_j\sigma_j^2|\mathbf{x}^{\mathbf{b}} \sim \text{Inv-}\chi^2\left(1, a_j\frac{2(n_j-1)s_j^2}{n_j}\right) = \text{Inv-}\chi^2\left(1,1\right), \tag{3.26}$$

for $j = 1, 2$, and 3. Thus, the multiplication by the tuning parameters results in equal automatic prior distributions for $a_1\sigma_1^2$, $a_2\sigma_2^2$, and $a_3\sigma_3^2$. Because these distributions are equal, all six possible adjusted orderings "$a_1\sigma_1^2 < a_2\sigma_2^2 < a_3\sigma_3^2$", ..., "$a_3\sigma_3^2 < a_2\sigma_2^2 < a_1\sigma_1^2$" are equally likely under the automatic prior. Therefore, the automatic prior probability of each adjusted ordering is equal to 1/6. Consequently, the Bayes factor in Equation (3.24) is equal to

$$B_{1u}^{aF} = 6 \times P\left(\sigma_1^2 < \sigma_2^2 < \sigma_3^2|\mathbf{x}\right). \tag{3.27}$$

Again, we consider data with $n_j = n = 20$ observations in each group with sample variances of $s_1^2 = 1$, $s_2^2 = s$, and $s_3^2 = s^2$, and we compute the aFBF for H_1 against H_u while letting s^2 increase from $\exp(-10) \approx 0$ to $\exp(10) \approx 22{,}000$. The results are shown in Figure 3.1. It can be seen that the aFBF (dashed line) converges to 6 as s^2 increases. This is a result of the posterior probability in Equation (3.27), which goes to 1 as s^2 increases, similar as in the FBF. Unlike in the FBF, however, the prior probability of the adjusted ordering is equal to 1/6. To give some more intuition, Figure 3.2b displays the adjusted parameter space when the sample variances are equal to $s_1^2 = 1$, $s_2^2 = 3$, and $s_3^2 = 9$, and the group sizes are equal to $n_j = n = 20$, for $j = 1, 2$, and 3. The plot illustrates how the adjusted parameter space adapts to the observed sample variances to ensure that the automatic prior probability of the adjusted ordering always equals 1/6. Because the aFBF for H_1 against H_u converges to 6, it can be argued that the order constrained hypothesis H_1 is 6 times more parsimonious than the unconstrained hypothesis.

Finally, note that in practice we do not recommend testing an inequality constrained hypothesis against the unconstrained hypothesis as in the above example. The reason is that the aFBF (and Bayes factors in general) is then bounded (e.g., by 6 in the case of $J = 3$ groups). This implies that we can never get decisive evidence in favor of H_1, even when observing very large effects in the direction of H_1 with very large samples. The main reason for testing H_1 against H_u in the above example was to illustrate how the parsimony of an inequality constrained hypothesis on variances is incorporated in the FBF and the aFBF. Generally, we would recommend testing an inequality constrained hypothesis H_1 against its complement H_2: not H_1 to avoid the issue of a bounded Bayes factor. For this test the aFBF would be equal to

$$B_{12}^{aF} = \frac{B_{1u}^{aF}}{B_{2u}^{aF}} = \frac{6 \times P\left(\sigma_1^2 < \sigma_2^2 < \sigma_3^2|\mathbf{x}\right)}{6/5 \times P\left(\text{not } \sigma_1^2 < \sigma_2^2 < \sigma_3^2|\mathbf{x}\right)} = 5 \times \frac{P\left(\sigma_1^2 < \sigma_2^2 < \sigma_3^2|\mathbf{x}\right)}{1 - P\left(\sigma_1^2 < \sigma_2^2 < \sigma_3^2|\mathbf{x}\right)}, \tag{3.28}$$

which does not have an upper bound.

3.4.5 Posterior Probabilities of the Hypotheses

When there are more than two hypotheses under investigation, it is useful to transform Bayes factors to posterior probabilities of the hypotheses. Here we show how to do this when working with the aFBF. To compute the posterior probabilities we first need to specify the prior probabilities of the hypotheses, denoted by $P(H_t)$, for $t = 1, \ldots, T$, where T is the number of hypotheses that are tested. These prior probabilities quantify how plausible each hypothesis is before observing the data. After observing the data, the prior probabilities can be updated using the marginal likelihoods from the aFBF in Equation (3.16) as follows:

$$P^{aF}(H_t|\mathbf{x}, \mathbf{b}) = \frac{m_t^{aF}(\mathbf{x}, \mathbf{b})P(H_t)}{\sum_{t'=1}^{T} m_{t'}^{aF}(\mathbf{x}, \mathbf{b})P(H_{t'})}. \tag{3.29}$$

The resulting posterior probabilities $P^{aF}(H_t|\mathbf{x}, \mathbf{b})$ quantify how plausible each hypothesis is after observing the data, for $t = 1, \ldots, T$. Note that the superscript aF is added to make it explicit that the posterior probabilities are computed using the marginal likelihoods based on the aFBF approach (see Equation (3.16)).

The default (or objective) choice in the literature is to set equal prior probabilities for the hypotheses, that is, $P(H_1) = \cdots = P(H_T) = 1/T$, which implies that it is assumed that all hypotheses are equally likely a priori (e.g., Berger & Mortera, 1999; Hoijtink, 2011; Mulder, Hoijtink, & de Leeuw, 2012). A consequence is that the ratio of the posterior probabilities of a pair of hypotheses is equal to the respective Bayes factor of these hypotheses. Because the prior probabilities sum to 1 (as well as the posterior probabilities), it is implicitly assumed that the true hypothesis is present in the set of constrained hypotheses under investigation. To ensure that this is the case it is recommended to always include the complement hypothesis when testing a set of constrained hypothesis on the variances. This was also done in the Math Garden example by including the complement hypothesis H_3 in Equation (3.4). Note that it is not recommended to set the prior probability of a hypothesis equal to the proportion of the unconstrained parameter space that it covers (e.g., 1/6 for $H_1\colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$). In that case the posterior probability of an inequality constrained hypothesis does not properly take the parsimony due to the inequality constraints into account (for details see Mulder, 2014a). Furthermore, the proportion of the unconstrained parameter space that is covered by a hypothesis involving at least one equality constraint is 0 (e.g., $H_2\colon \sigma_1^2 = \sigma_2^2 < \sigma_3^2$ describes a plane in the unconstrained space, which has a volume of 0). However, a prior probability of 0 results in a posterior probability of 0 (see Equation (3.29)), which means that there can never be evidence in favor of an equality constrained hypothesis.

3.5 Simulation Study: Performance of the Adjusted Fractional Bayes Factor

The goal of our simulation study is to assess the performance of the adjusted fractional Bayes factor when testing equality and inequality constrained hypotheses on variances. Our focus is both on consistency (i.e., does the aFBF select the true hypothesis when the sample size is large) and small-sample performance.

3.5.1 Design

The performance of the aFBF is examined as a function of the following four factors:

1. *Number of groups:* We compared variances of $J = 3$ and 5 groups.

2. *Population:* For each of the two numbers of groups we considered five populations differing in the structure of the population variances. An overview is given in Table 3.2. The first was a null population in which all population variances were equal, $\sigma_1^2 = \cdots = \sigma_J^2$. The second population was one in which the variances followed the hypothesized order $\sigma_1^2 < \cdots < \sigma_J^2$. We refer to this population as the order population. The mixed population featured equalities as well as inequalities among the variances. For $J = 3$ groups the structure of the population variances was $\sigma_1^2 < \sigma_2^2 = \sigma_3^2$, whereas for $J = 5$ groups it was $\sigma_1^2 < \sigma_2^2 = \sigma_3^2 < \sigma_4^2 = \sigma_5^2$. The near order population was identical to the order population with the exception that the order of the two groups with the largest variances was reversed, $\sigma_1^2 < \cdots < \sigma_J^2 < \sigma_{J-1}^2$. Finally, in the reverse order population the order of the population variances was the opposite of that in the order population, $\sigma_J^2 < \cdots < \sigma_1^2$. Note that the reverse order is maximally different from the hypothesized order. We included the near order and the reverse order population to check how much data is needed to detect that the hypothesized order is slightly different from the true order (near order population) or very different from the true order (reverse order population).

3. *Effect size:* In all populations except the null population we considered three effect sizes: small, medium, and large. The effect size is given by the ratio of the largest population variance to the smallest population variance. To our knowledge no guidelines exist as to what population variance ratios constitute a small, medium, and large effect. We therefore based our effect sizes on well-known guidelines for testing equality of means of two independent populations. These guidelines state that the power to detect a small, medium, and large effect equals 0.8 for $\alpha = 0.05$ and sample sizes of 310, 51, and 21 in each group, respectively (Faul, Erdfelder, Buchner, & Lang, 2009). We used these numbers to determine the population variances in our simulation study in four steps: First, we used the sample sizes of 310, 51, 21 to determine the noncentrality parameter λ of the non-central F-distribution such that the power for testing equality of variances of two independent populations equals 0.8. For a small,

Table 3.2: Structure of population variances in five populations for $J \in \{3, 5\}$ groups.

Population	$J = 3$	$J = 5$
Null	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2$	$\sigma_1^2 = \cdots = \sigma_5^2$
Order	$\sigma_1^2 < \sigma_2^2 < \sigma_3^2$	$\sigma_1^2 < \cdots < \sigma_5^2$
Mixed	$\sigma_1^2 < \sigma_2^2 = \sigma_3^2$	$\sigma_1^2 < \sigma_2^2 = \sigma_3^2 < \sigma_4^2 = \sigma_5^2$
Near order	$\sigma_1^2 < \sigma_3^2 < \sigma_2^2$	$\sigma_1^2 < \cdots < \sigma_5^2 < \sigma_4^2$
Reverse order	$\sigma_3^2 < \sigma_2^2 < \sigma_1^2$	$\sigma_5^2 < \cdots < \sigma_1^2$

Table 3.3: Population variances in the simulation study.

		$J = 3$			$J = 5$				
Population	Effect	σ_1^2	σ_2^2	σ_3^2	σ_1^2	σ_2^2	σ_3^2	σ_4^2	σ_5^2
Null	No	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Order	Small	1.00	1.20	1.43	1.00	1.11	1.23	1.37	1.52
	Medium	1.00	1.53	2.33	1.00	1.27	1.61	2.05	2.60
	Large	1.00	1.89	3.59	1.00	1.42	2.03	2.88	4.10
Mixed	Small	1.00	1.33	1.33	1.00	1.20	1.20	1.43	1.43
	Medium	1.00	2.00	2.00	1.00	1.53	1.53	2.33	2.33
	Large	1.00	2.94	2.94	1.00	1.89	1.89	3.59	3.59
Near order	Small	1.00	1.43	1.20	1.00	1.11	1.23	1.52	1.37
	Medium	1.00	2.33	1.53	1.00	1.27	1.61	2.60	2.05
	Large	1.00	3.59	1.89	1.00	1.42	2.03	4.10	2.88
Reverse order	Small	1.43	1.20	1.00	1.52	1.37	1.23	1.11	1.00
	Medium	2.33	1.53	1.00	2.60	2.05	1.61	1.27	1.00
	Large	3.59	1.89	1.00	4.10	2.88	2.03	1.42	1.00

medium, and large effect, we obtain values of λ of 100.74, 49.94, and 38.80, respectively. Second, we computed the population variance ratio as VR $= (n - 1 + \lambda)/(n - 1)$, which equals the expected value of the non-central F-distribution. Here the common sample size n equals 310, 51, and 21 if λ equals 100.74, 49.94, and 38.80, respectively. The resulting ratios are 1.33, 2.00, and 2.94 for a small, medium, and large effect, respectively. Third, to determine σ_J^2/σ_1^2 for $J = 3$ and 5 groups, we computed the $(J-1)/J$ quantile of a uniform distribution with minimum value 1 and maximum value $2 \times \text{VR} - 1$. This results in population variance ratios that increase with the number of groups J, which is supported by empirical findings (see, e.g., Ruscio & Roche, 2012). In all populations we set $\sigma_1^2 = 1$, so that σ_J^2 is determined by the population variance ratio. Fourth, we computed the intermediate population variances as $\sigma_j^2 = \left(\sigma_J^2\right)^{(j-1)/(J-1)}$ for $j = 2, \ldots, J-1$. As a result, the ratio of adjacent population variances is constant, that is, $\sigma_2^2/\sigma_1^2 = \cdots = \sigma_J^2/\sigma_{J-1}^2$. Table 3.3 gives an overview of all population variances used in the simulation study. Note that in the mixed population with $J = 3$ groups we used the population variance ratios from the $J = 2$ groups case, that is, 1.33, 2.00, and 2.94. We did so because, in fact, there are only two distinct variances in this population (cf. Table 3.2). Similarly, in the mixed population with $J = 5$ groups we used the population variance ratios from the $J = 3$ groups case.

4. *Sample size:* We used a balanced design with common sample sizes of 5, 10, 20, 50, 100, 200, 500, 1,000, 2,000, and 5,000.

Thus, in total there were 260 conditions, 2 (number of groups) $\times$ 10 (sample size) = 20 for the null population and 2 (number of groups) $\times$ 4 (population) $\times$ 3 (effect size) $\times$

10 (sample size) = 240 for the remaining four populations.

3.5.2 Hypotheses and Data Generation

In each of the five populations we tested three hypotheses. An overview is given in Table 3.4. In the null population, the order population, the near order population, and the reverse order population we tested the following three hypotheses: $H_0\colon \sigma_1^2 = \cdots = \sigma_J^2$, $H_1\colon \sigma_1^2 < \cdots < \sigma_J^2$, and $H_2\colon$ not $\sigma_1^2 < \cdots < \sigma_J^2$. Note that $H_2\colon$ not $\sigma_1^2 < \cdots < \sigma_J^2$ is equivalent to $H_2\colon$ not (H_0 or H_1) because the probability of the event that the variances are exactly equal is 0 under the unconstrained hypothesis. In Table 3.4 the true hypothesis (i.e., the hypothesis that correctly describes the structure of the population variances) is flagged with an asterisk (H^*). Note that for the near order and the reverse order population the true hypothesis is contained in the complement H_2. In the mixed population with $J = 3$ groups we tested $H_0\colon \sigma_1^2 = \sigma_2^2 = \sigma_3^2$, $H_1\colon \sigma_1^2 < \sigma_2^2 = \sigma_3^2$, and $H_2\colon \sigma_1^2 < \left(\sigma_2^2, \sigma_3^2\right)$. Here H_2 states that the variances in Groups 2 and 3 are larger than in Group 1, but not necessarily equal. In the mixed population with $J = 5$ groups we tested the corresponding hypotheses $H_0\colon \sigma_1^2 = \cdots = \sigma_5^2$, $H_1\colon \sigma_1^2 < \sigma_2^2 = \sigma_3^2 < \sigma_4^2 = \sigma_5^2$, and $H_2\colon \sigma_1^2 < \left(\sigma_2^2, \sigma_3^2\right) < \left(\sigma_4^2, \sigma_5^2\right)$.

In each of the 260 conditions we generated 1,000 data sets. The population variances were specified according to Table 3.3. In all conditions we set $\mu_1 = \cdots = \mu_J = 0$. We may do so because the aFBF is independent of the population means (in Equations (3.16) and (3.17) it can be seen that the marginal likelihood in the aFBF approach does not depend on the sample means). For each of the 1,000 data sets we computed the evidence in favor of the true hypothesis. We used two measures of evidence: The first is the logarithm of the Bayes factor in favor of the true hypothesis H_t, $\log\left(B_{tt'}^{aF}\right)$. The second measure is the posterior probability of the true hypothesis, $P^{aF}(H_t|\mathbf{x}, \mathbf{b})$, which was computed assuming equal prior probabilities of the hypotheses. The log Bayes factors and posterior probabilities were computed using minimal fractions of $b_j = 2/n_j$, for $j = 1, \ldots, J$. Eventually, we computed the median of the 1,000 log Bayes factors and posterior probabilities.

Table 3.4: Hypotheses tested in the simulation study. In each population we tested three hypotheses. The true hypothesis is flagged with an asterisk (H^*). Here $J \in \{3, 5\}$ indicates the number of groups.

Population	Tested hypotheses		
Null	$H_0^*\colon \sigma_1^2 = \cdots = \sigma_J^2$	$H_1\colon \sigma_1^2 < \cdots < \sigma_J^2$	$H_2\colon$ not $\sigma_1^2 < \cdots < \sigma_J^2$
Order	$H_0\colon \sigma_1^2 = \cdots = \sigma_J^2$	$H_1^*\colon \sigma_1^2 < \cdots < \sigma_J^2$	$H_2\colon$ not $\sigma_1^2 < \cdots < \sigma_J^2$
Mixed, $J = 3$	$H_0\colon \sigma_1^2 = \sigma_2^2 = \sigma_3^2$	$H_1^*\colon \sigma_1^2 < \sigma_2^2 = \sigma_3^2$	$H_2\colon \sigma_1^2 < \left(\sigma_2^2, \sigma_3^2\right)$
Mixed, $J = 5$	$H_0\colon \sigma_1^2 = \cdots = \sigma_5^2$	$H_1^*\colon \sigma_1^2 < \sigma_2^2 = \sigma_3^2 < \sigma_4^2 = \sigma_5^2$	$H_2\colon \sigma_1^2 < \left(\sigma_2^2, \sigma_3^2\right) < \left(\sigma_4^2, \sigma_5^2\right)$
Near order	$H_0\colon \sigma_1^2 = \cdots = \sigma_J^2$	$H_1\colon \sigma_1^2 < \cdots < \sigma_J^2$	$H_2^*\colon$ not $\sigma_1^2 < \cdots < \sigma_J^2$
Reverse order	$H_0\colon \sigma_1^2 = \cdots = \sigma_J^2$	$H_1\colon \sigma_1^2 < \cdots < \sigma_J^2$	$H_2^*\colon$ not $\sigma_1^2 < \cdots < \sigma_J^2$

3.5.3 Results

The results of the simulation study are shown in Figures 3.3 to 3.7. Each figure shows the results for one of the five populations we considered. The plots show the median log Bayes factors in favor of the true hypothesis (left-hand column) and the median posterior probability of the true hypothesis (flagged with an asterisk in Table 3.4; right-hand column) for $J = 3$ groups (top row) and $J = 5$ groups (bottom row) as a function of the common sample size $n_1 = \cdots = n_J = n$. For the null population the results for $J = 3$ and $J = 5$ groups are combined in one pair of plots, see Figure 3.3. Two important general conclusions can be drawn from the figures. First, the aFBF is consistent. For all numbers of groups, populations, and effect sizes the posterior probability of the true hypothesis was equal or close to 1 for a common sample size of 5,000. Second, the performance of the aFBF was similar for $J = 3$ and $J = 5$ groups, with the relevant differences being that for $J = 5$ groups the null hypothesis received stronger support and larger sample sizes were needed to reject a false null hypothesis. We now focus on small-sample performance of the aFBF for each population separately.

Null Population

Figure 3.3 shows the simulation results for the null population. The plots show that the evidence in favor of the true hypothesis H_0 increased with sample size. The log Bayes factor $\log\left(B_{01}^{aF}\right)$ was consistently larger than $\log\left(B_{02}^{aF}\right)$ because under H_0 the

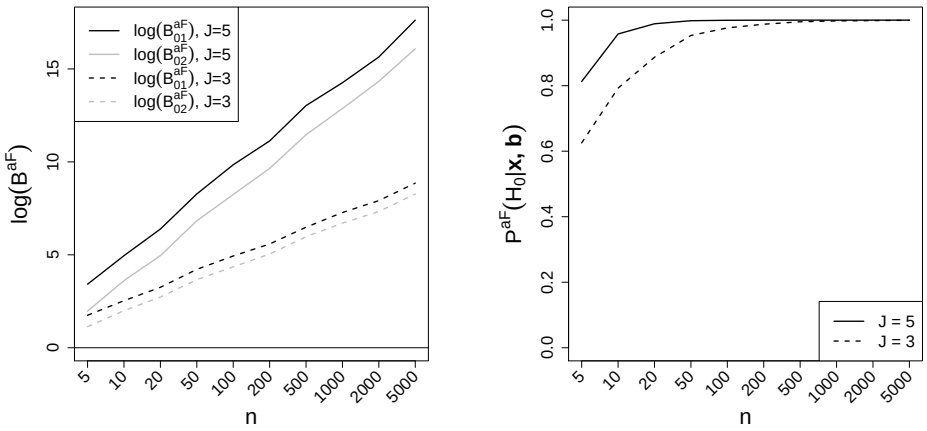


Figure 3.3: Simulation results for a null population in which all population variances were equal, $\sigma_1^2 = \cdots = \sigma_J^2$, for $J = 3$ groups (dashed lines) and $J = 5$ groups (solid lines). We tested the true hypothesis $H_0 \colon \sigma_1^2 = \cdots = \sigma_J^2$ against the two competing hypotheses $H_1 \colon \sigma_1^2 < \cdots < \sigma_J^2$ and $H_2 \colon$ not $\sigma_1^2 < \cdots < \sigma_J^2$. The plots show the median log Bayes factors (left-hand plot) testing H_0 against H_1 (black lines) and H_0 against H_2 (gray lines) and the median posterior probability of H_0 (right-hand plot) as a function of the common sample size $n_1 = \cdots = n_J = n$.

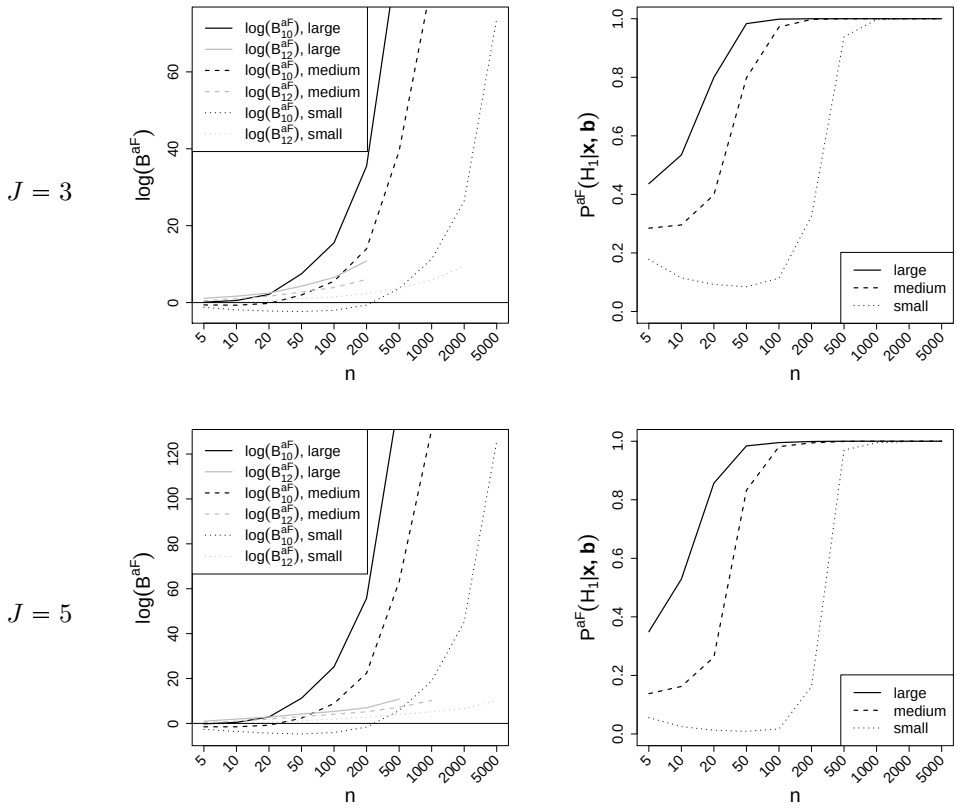


Figure 3.4: Simulation results for an order population in which the structure of the population variances was $\sigma_1^2 < \cdots < \sigma_J^2$, with $J \in \{3, 5\}$. We considered three effect sizes: small (dotted lines), medium (dashed lines), and large (solid lines). We tested the true hypothesis $H_1 \colon \sigma_1^2 < \cdots < \sigma_J^2$ against the two competing hypotheses $H_0 \colon \sigma_1^2 = \cdots = \sigma_J^2$ and $H_2 \colon$ not $\sigma_1^2 < \cdots < \sigma_J^2$. The plots show the median log Bayes factors (left-hand column) testing H_1 against H_0 (black lines) and H_1 against H_2 (gray lines) and the median posterior probability of H_1 (right-hand column) as a function of the common sample size $n_1 = \cdots = n_J = n$. In the log Bayes factors plots the gray lines are discontinued due to numerical reasons (see text).

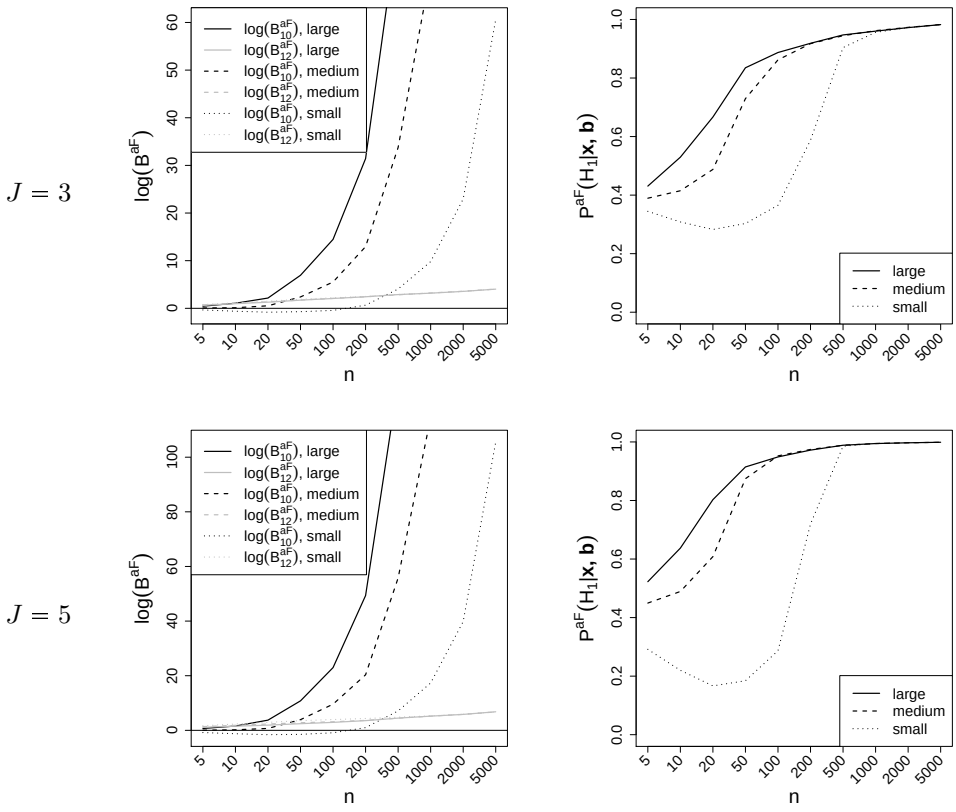


Figure 3.5: Simulation results for a mixed population. For $J = 3$ groups the structure of the population variances was $\sigma_1^2 < \sigma_2^2 = \sigma_3^2$, whereas for $J = 5$ groups it was $\sigma_1^2 < \sigma_2^2 = \sigma_3^2 < \sigma_4^2 = \sigma_5^2$. We considered three effect sizes: small (dotted lines), medium (dashed lines), and large (solid lines). For $J = 3$ groups we tested the true hypothesis $H_1 \colon \sigma_1^2 < \sigma_2^2 = \sigma_3^2$ against the two competing hypotheses $H_0 \colon \sigma_1^2 = \sigma_2^2 = \sigma_3^2$ and $H_2 \colon \sigma_1^2 < \left(\sigma_2^2, \sigma_3^2\right)$. For $J = 5$ groups we tested the true hypothesis $H_1 \colon \sigma_1^2 < \sigma_2^2 = \sigma_3^2 < \sigma_4^2 = \sigma_5^2$ against $H_0 \colon \sigma_1^2 = \cdots = \sigma_5^2$ and $H_2 \colon \sigma_1^2 < \left(\sigma_2^2, \sigma_3^2\right) < \left(\sigma_4^2, \sigma_5^2\right)$. The plots show the median log Bayes factors (left-hand column) testing H_1 against H_0 (black lines) and H_1 against H_2 (gray lines) and the median posterior probability of H_1 (right-hand column) as a function of the common sample size $n_1 = \cdots = n_J = n$.

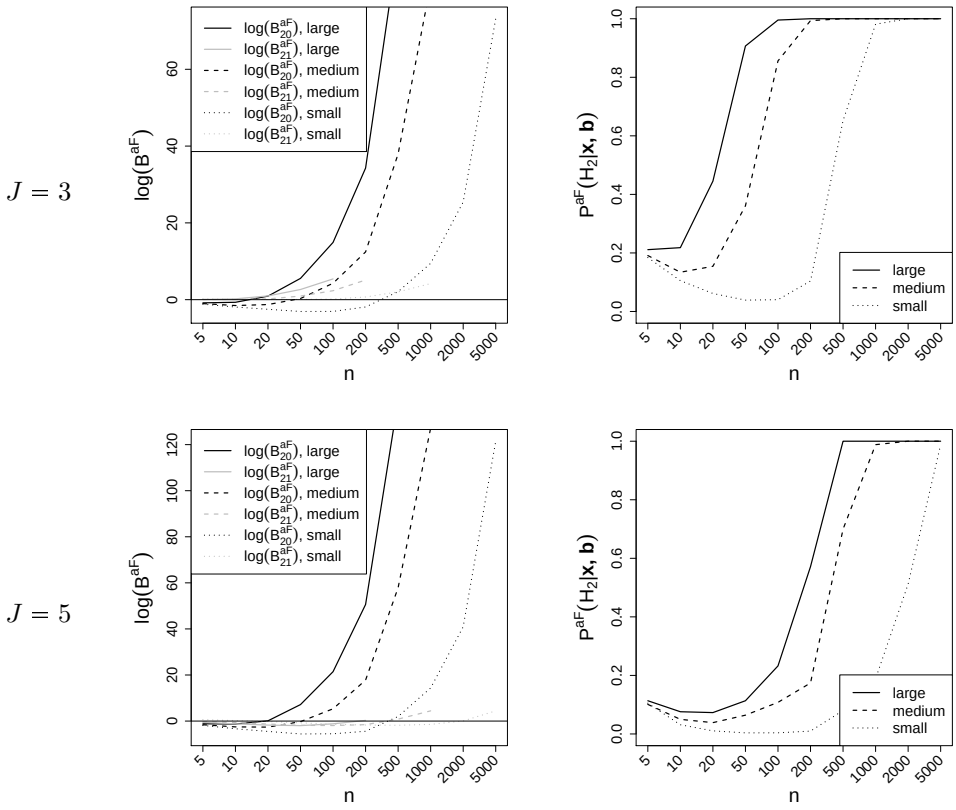


Figure 3.6: Simulation results for a near order population in which the structure of the population variances was $\sigma_1^2 < \cdots < \sigma_J^2 < \sigma_{J-1}^2$, with $J \in \{3, 5\}$. We considered three effect sizes: small (dotted lines), medium (dashed lines), and large (solid lines). We tested three hypotheses: H_0: $\sigma_1^2 = \cdots = \sigma_J^2$, H_1: $\sigma_1^2 < \cdots < \sigma_J^2$, and H_2: not $\sigma_1^2 < \cdots < \sigma_J^2$. Note that the true hypothesis is the complement H_2. The plots show the median log Bayes factors (left-hand column) testing H_2 against H_0 (black lines) and H_2 against H_1 (gray lines) and the median posterior probability of H_2 (right-hand column) as a function of the common sample size $n_1 = \cdots = n_J = n$. In the log Bayes factors plots the gray lines are discontinued due to numerical reasons (see text).

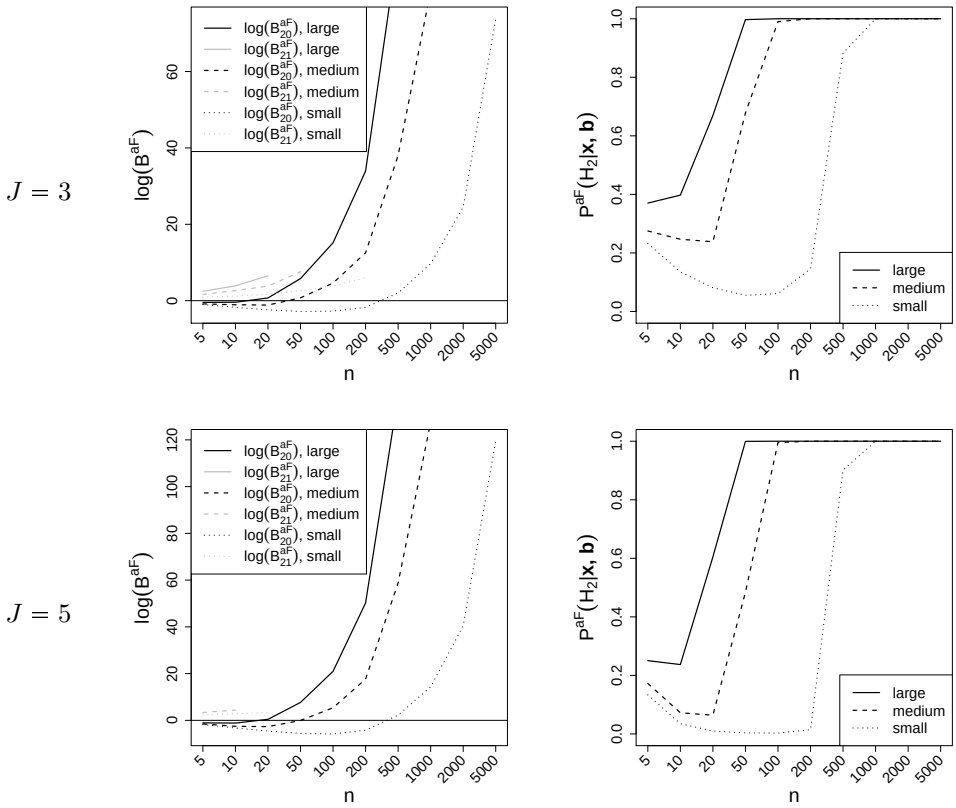


Figure 3.7: Simulation results for a reverse order population in which the structure of the population variances was $\sigma_J^2 < \cdots < \sigma_1^2$, with $J \in \{3, 5\}$. We considered three effect sizes: small (dotted lines), medium (dashed lines), and large (solid lines). We tested three hypotheses: $H_0 \colon \sigma_1^2 = \cdots = \sigma_J^2$, $H_1 \colon \sigma_1^2 < \cdots < \sigma_J^2$, and $H_2 \colon$ not $\sigma_1^2 < \cdots < \sigma_J^2$. Note that the true hypothesis is the complement H_2. The plots show the median log Bayes factors (left-hand column) testing H_2 against H_0 (black lines) and H_2 against H_1 (gray lines) and the median posterior probability of H_2 (right-hand column) as a function of the common sample size $n_1 = \cdots = n_J = n$. In the log Bayes factors plots the gray lines are discontinued due to numerical reasons (see text).

order constrained hypotheses H_1 fits worse than the complement H_2. This is because H_1 is more restrictive than H_2. In the right-hand plot we see that for samples of size $n = 5$ the posterior probability of H_0 was greater than 0.6, and samples as small as $n = 10$ yielded a posterior probability of about 0.8. The probability is so high even for small samples because neither H_1 nor H_2 are good competitors to H_0, particularly so for $J = 5$ groups.

Order Population

Figure 3.4 shows the simulation results for the order population. The plots illustrate that the evidence in favor of the true hypothesis H_1 did not increase with sample size when the effect was small. This is a consequence of the fact that small effects can be better explained by the null hypothesis than by the order constrained hypothesis when the sample size is small. The posterior probability of the true hypothesis H_1 was at least 0.8 for sample sizes of about 500 (small effect), 100 (medium effect), and 50 (large effect), respectively. Finally, note that the gray lines for the log Bayes factor $\log\left(B_{12}^{aF}\right)$ are discontinued at some point. This is due to numerical reasons: In the computation of the discontinued log Bayes factors we had to divide by the posterior probability that the inequality constraints do not hold. This was estimated by the proportion of draws from the unconstrained posterior distribution for which these constraints do not hold. For large samples this proportion was often 0, so that the corresponding log Bayes factor was undefined. If this happened for the majority of the 1,000 replications in the simulation, then the median log Bayes factor was undefined as well. Note that theoretically the discontinued log Bayes factors keep increasing because the posterior probability that the inequality constraints do not hold approaches 0 as the sample size increases.

Mixed Population

The results for the mixed population are shown in Figure 3.5. Similar to the order population, the evidence in favor of the true hypothesis did not increase with sample size when the effect was small. For $J = 3$ groups and a small effect the evidence only increased for sample sizes larger than 50. Actually, in this case the log Bayes factor $\log\left(B_{10}^{aF}\right)$ favored the null hypothesis until the sample size surpassed the $n = 200$ mark. The reason is the same as for the order population, namely, that small effects can be better explained by the null hypothesis when the sample size is small. The posterior probability was above 0.8 for samples of size 500 (small effect), 50 to 100 (medium effect), and 50 (large effect), respectively. Note that it approached 1 somewhat more slowly than in the order population. This is due to the similarity of H_1 and H_2. Finally, note that the log Bayes factor $\log\left(B_{12}^{aF}\right)$ did not depend on the effect size. It was approximately the same under all effects, which can be seen from the three gray lines overlapping. This is because H_1 and H_2 essentially state the same effect, namely, that σ_1^2 is smaller than σ_2^2 and σ_3^2.

Near Order Population

Figure 3.6 shows the simulation results for the near order population. Again, the evidence in favor of the true hypothesis did not generally increase with sample size.

For a small effect it only increased for sample sizes larger than about 100. For $J = 3$ groups the posterior probability of the true hypothesis reached values of at least 0.8 for samples of size 1,000 (small effect), 100 (medium effect), and 50 (large effect), respectively. For $J = 5$ groups substantially larger samples were required (5,000, 1,000, and 200, respectively). This is mainly because the ratio of adjacent population variances is smaller in the $J = 5$ groups case (see Table 3.3), which makes it more difficult to detect that the two largest population variances are ordered as $\sigma^2_J < \sigma^2_{J-1}$ instead of $\sigma^2_{J-1} < \sigma^2_J$. Similar to the order population, the log Bayes factor $\log\left(B^{aF}_{21}\right)$ (gray lines) could not be computed for larger sample sizes due to numerical reasons.

Reverse Order Population

The evidence in favor of the true hypothesis did not generally increase with sample size, see Figure 3.7. For instance, for a small effect and $J = 5$ groups the evidence only increased for sample sizes larger than 200. The evidence in favor of the true hypothesis increased faster for the reverse order than for the near order population because the reverse order population is less in agreement with the order constrained hypothesis H_1 than the near order population. The posterior probability of the true hypothesis was greater than 0.8 for sample sizes of 500 (small effect), 100 (medium effect), and 50 (large effect), respectively. Again, we see discontinued log Bayes factors due to numerical reasons.

3.5.4 Conclusion

In conclusion, the results of the simulation show that the aFBF performed well in all five populations we considered. In particular, the results indicate that the aFBF is consistent in the sense that it selects the true hypothesis if the sample size is large enough. Naturally, for small effects we needed larger samples to detect the true hypothesis than for large effects.

We also performed the simulation with unequal group sizes to check for robustness of the results obtained with equal group sizes. All settings except the sample sizes were identical to the simulation with equal group sizes. We provide the sample sizes and results of the simulation with unequal group sizes in the supplemental material in Appendix 3.D. The results confirm the findings from the simulation with equal group sizes discussed above.

3.6 Illustrative Example: The Math Garden (Continued)

After logging into the Math Garden, children are directed to a page showing a garden in which plants represent games covering different domains of mathematics, see Figure 3.8a. In this illustrative example we focus on the four most played games: addition, subtraction, multiplication, and division. Each of these games consists of over 700 items ranging from easy (e.g., $2 + 2$) to difficult (e.g., $340 + 87$). Figure 3.8b shows an exemplary addition item. By clicking on a plant the player starts a session of 15 items. The items are adaptively selected based on a player's ability. The system takes both

(a) The garden page (b) An addition item

Figure 3.8: Two screenshots of the Math Garden. Figure (a) shows the garden page where each plant represents a game measuring a different aspect of mathematics. Figure (b) shows an exemplary addition item.

accuracy of responses and response times into account to estimate a player's ability. For details on the Math Garden and the underlying IRT model we refer the interested reader to Klinkenberg et al. (2011) and Maris and van der Maas (2012).

We used two criteria for extracting ability estimates from the Math Garden database. The first criterion concerns the grade a student is in. Aunola et al. (2004) hypothesize that systematic instruction at school functions as a sort of treatment that results in an increase or a decrease of the variability of abilities across grades. It thus makes sense to only consider grades in which the treatment is administered to the students. In the Netherlands children are taught addition and subtraction at school from Grade 1 through Grade 5. For the addition and the subtraction domain we therefore extracted ability estimates of students in Grades 1 through 5. Multiplication and division is taught from Grade 3 through 6, which is why for these two domains we extracted ability estimates of students in these grades. The second criterion we used is that children have to have played at least 45 items (i.e., three sessions) in the week prior to extraction. The reason for this is twofold. First, the more items a student plays the more precise their ability can be estimated. Experience has shown that after 45 items ability estimates are reasonably precise and stable. Second, we require children to have played the items in one week in order to avoid that there is too much learning going on due to treatment at school.

Table 3.5 shows the sample size and sample variance for each grade and mathematical domain. We use the symbols $+$, $-$, $\times$, and $\div$ to refer to the corresponding game in the Math Garden. Furthermore, the table shows the variance ratio, which is given by the ratio of a sample variance to the smallest sample variance in the corresponding domain. In the addition and the subtraction domain it can be seen that the sample variances do not follow an increasing order. The variance decreases from Grade 1 to Grade 2, and subsequently increases from Grade 2 to Grade 5. In the multiplication and the division domain, however, the sample variances follow an increasing order from Grade 3 to Grade 6.

Table 3.6 shows the posterior probability of the hypotheses $H_0\colon \sigma_1^2 = \cdots = \sigma_J^2$, $H_1\colon \sigma_1^2 < \cdots < \sigma_J^2$, $H_2\colon \sigma_J^2 < \cdots < \sigma_1^2$, and $H_3\colon$ not (H_0 or H_1 or H_2) for each domain. We computed the posterior probabilities assuming equal prior probabilities

Table 3.5: Descriptive statistics for the Math Garden data. The symbols +, −, ×, and ÷ refer to the corresponding domain in the Math Garden. The variance ratio is the ratio of the sample variance to the smallest sample variance in the corresponding domain.

	Sample size				Sample variance				Variance ratio			
Grade	+	−	×	÷	+	−	×	÷	+	−	×	÷
1	4,336	1,471	—	—	7.22	7.45	—	—	1.25	1.17	—	—
2	4,080	2,663	—	—	5.76	6.35	—	—	1.00	1.00	—	—
3	2,396	1,763	3,567	1,434	7.26	9.76	4.69	24.20	1.26	1.54	1.00	1.00
4	1,551	1,123	2,968	1,907	9.86	13.83	8.04	27.10	1.71	2.18	1.71	1.12
5	1,239	756	2,197	1,815	14.57	16.69	12.99	33.99	2.53	2.63	2.77	1.40
6	—	—	1,094	1,117	—	—	20.64	45.65	—	—	4.40	1.89

of the hypotheses. The posterior probabilities are (close to) 0.00 or 1.00 due to the large sample sizes in combination with the considerable effect sizes (cf. the results of the simulation study). One immediate conclusion we can draw is that there is no evidence in favor of either H_0 or H_2 in any of the domains, as can be seen from their posterior probabilities being 0.00. The hypotheses of equality of variances and decreasing variances can therefore safely be rejected. Furthermore, in the addition and the subtraction domain we can rule out H_1 given posterior probabilities of 0.00 and 0.03, respectively. The decrease in variance from Grade 1 to Grade 2 in combination with the large sample sizes makes an increasing order of the variances highly unlikely. We conclude that in the addition and the subtraction domain something other than H_0, H_1, and H_2 is going on, as is indicated by the posterior probabilities of the complement H_3 being 1.00 and 0.97, respectively. In the multiplication and the division domain, however, there is very strong evidence in favor of an increase in variance, with posterior probabilities of H_1 of 1.00. In these domains we can rule out H_0, H_2, and H_3, as is indicated by posterior probabilities of these hypotheses of 0.00.

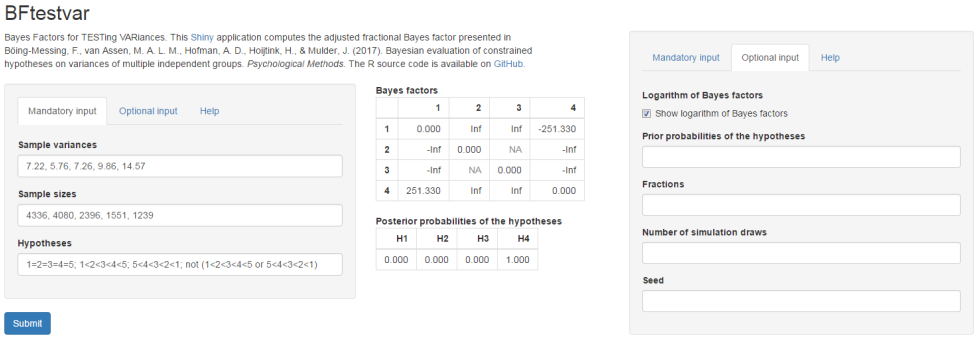
Table 3.6: Results of the analysis of the Math Garden data. The symbols +, −, ×, and ÷ refer to the corresponding domain in the Math Garden.

Result	+	−	×	÷
$P^{aF}(H_0\colon \sigma_1^2 = \cdots = \sigma_J^2 \mid \mathbf{x}, \mathbf{b})$	0.00	0.00	0.00	0.00
$P^{aF}(H_1\colon \sigma_1^2 < \cdots < \sigma_J^2 \mid \mathbf{x}, \mathbf{b})$	0.00	0.03	1.00	1.00
$P^{aF}(H_2\colon \sigma_J^2 < \cdots < \sigma_1^2 \mid \mathbf{x}, \mathbf{b})$	0.00	0.00	0.00	0.00
$P^{aF}(H_3\colon \text{not } (H_0 \text{ or } H_1 \text{ or } H_2) \mid \mathbf{x}, \mathbf{b})$	1.00	0.97	0.00	0.00

3.7 Software Application for Computing the Adjusted Fractional Bayes Factor

We provide a Shiny application for computing the adjusted fractional Bayes factor. Shiny (Chang, Cheng, Allaire, Xie, & McPherson, 2015) is a framework for creating interactive applications using the R language for statistical computing (R Core Team, 2015). The advantage of Shiny applications is that the user does not need to read or write R code.

Figure 3.9 shows two screenshots of our Shiny application. On the left-hand side of Figure 3.9a one can see the "Mandatory input" tab panel. Here the user needs to specify the sample variances, sample sizes, and hypotheses. The screenshot shows the input for the addition domain in the Math Garden example. As can be seen, the hypotheses need to be specified using group numbers $1, \ldots, J$. For example, the hypothesis $H_1 \colon \sigma_1^2 < \cdots < \sigma_5^2$ from the Math Garden example is specified as "$1 < 2 < 3 < 4 < 5$". Note that inequality constraints need to be specified using the less-than symbol ($<$); the greater-than symbol ($>$) is not supported. The complement of an order constrained hypothesis can be specified using the string "not" in the beginning (e.g., "not $1 < 2 < 3 < 4 < 5$"). Note that the complement of a hypothesis containing at least one equality constraint is equivalent to the unconstrained hypothesis. This is because the probability of the event that two or more variances are exactly equal is 0 under the unconstrained hypothesis. For example, the hypothesis $H_1 \colon \sigma_1^2 = \sigma_2^2 < \sigma_3^2$ describes a plane in the unconstrained space, which has a probability of 0 (in the sense that the volume is 0). The complement $H_2 \colon$ not $\sigma_1^2 = \sigma_2^2 < \sigma_3^2$ comprises the entire space except the plane in H_1, which is mathematically equivalent to the unconstrained space. Hence the complement H_2 is equivalent to the unconstrained hypothesis $H_u \colon \sigma_1^2, \sigma_2^2, \sigma_3^2$. For the same reason the hypothesis $H_3 \colon$ not $\left(\sigma_1^2 = \cdots = \sigma_5^2 \text{ or } \sigma_1^2 < \cdots < \sigma_5^2 \text{ or } \sigma_5^2 < \cdots < \sigma_1^2\right)$ from the Math Garden example is specified as "not $(1 < 2 < 3 < 4 < 5$ or $5 < 4 < 3 < 2 < 1)$" in the



(a) Mandatory input and output (b) Optional input

Figure 3.9: Two screenshots of the Shiny application for computing the adjusted fractional Bayes factor. Figure (a) shows the "Mandatory input" tab panel and the output (Bayes factors and posterior probabilities of the hypotheses). Figure (b) shows the "Optional input" tab panel.

application (see Figure 3.9a).

Figure 3.9b shows the "Optional input" tab panel. Here the user may specify more advanced settings. Using the checkbox one can control whether the application shows Bayes factors or log Bayes factors (the latter is sometimes also referred to as the weight of evidence). In the next field the user may specify prior probabilities of the hypotheses. By default (i.e., if the field is empty) the posterior probabilities of the hypotheses are computed assuming equal prior probabilities. The "Fractions" field can be used to specify custom fractions $b_1, \ldots, b_J$. If the field is empty, the application uses the minimal information approach and sets $b_j = 2/n_j$ by default. Computing the marginal likelihood under an inequality constrained hypothesis involves sampling from the posterior and the prior distribution of the group variances. In the field "Number of simulation draws" one may specify how often to draw from the posterior and the prior. By default the application simulates 100,000 draws. In the last field the user may specify a custom seed in order to reproduce results exactly in the case of testing inequality constrained hypotheses (which requires simulating from the posterior and the prior). The "Help" tab panel contains detailed instructions on how to use the application.

Once all input has been specified, clicking on the "Submit" button initiates the computation of the results. Computation time mostly depends on the number of simulation draws, the number of hypotheses, and the number of inequality constraints. For example, the analysis shown in the screenshots should be completed within a few seconds. The results are shown in the output on the right-hand side of Figure 3.9a. The output consists of two tables, one showing the (log) Bayes factors and one showing the posterior probabilities of the hypotheses. The screenshot shows the results for the addition domain in the Math Garden example. In the "Bayes factors" table, the cell in row $t \in \{1, 2, 3, 4\}$ and column $t' \in \{1, 2, 3, 4\}$ contains the logarithm of the Bayes factor $B^{aF}_{tt'}$ (because we ticked the "Show logarithm of Bayes factors" checkbox in the optional input, see Figure 3.9b). For example, the cell in row 4 and column 1 contains the logarithm of the Bayes factor B^{aF}_{41} testing H_4: not $\left(\sigma_1^2 < \cdots < \sigma_5^2 \text{ or } \sigma_5^2 < \cdots < \sigma_1^2\right)$ against $H_1 : \sigma_1^2 = \cdots = \sigma_5^2$ (note that the hypotheses are numbered consecutively starting with 1). The log Bayes factor equals 251.33, which means that the evidence in the data in favor of H_4 is exp(251.33) times as strong as the evidence in favor of H_1. Some log Bayes factors are infinite because the marginal likelihoods under $H_2 : \sigma_1^2 < \cdots < \sigma_5^2$ and $H_3 : \sigma_5^2 < \cdots < \sigma_1^2$ are approximated as 0. As a result, the logarithms of the Bayes factors B^{aF}_{23} and B^{aF}_{32} are undefined, which is why in the corresponding cells in the table it says NA (for "not available").

To run our Shiny application follow these six steps:

1. Download and install R from https://cran.r-project.org/.

2. Launch R.

3. Copy the following R code and paste it into the R console:

   ```
   install.packages("shiny")
   ```

   Hit the Enter key and select a mirror. This will install the Shiny package.

4. Copy the following R code and paste it into the R console:

   ```
   library(shiny)
   ```

   Hit the Enter key. This will load the Shiny package.

5. Copy the following R code and paste it into the R console:

   ```
   runGitHub("BFtestvar", "fboeingmessing")
   ```

   Hit the Enter key. The Shiny application will open in your browser.

6. When you have completed your analyses you need to stop the application in order to be able to close R. To do so click on the red "STOP" button in the R menu bar.

Note that steps 1 and 3 only need to be performed the first time you use the application. The R source code of the application is available at https://github.com/fboeingmessing/BFtestvar.

3.8 Discussion

In this chapter we developed a Bayes factor for testing equality and inequality constrained hypotheses on variances. Our method is based on an adjustment of the fractional Bayes factor (O'Hagan, 1995) such that it properly incorporates the parsimony of inequality constrained hypotheses. Using our adjusted fractional Bayes factor we can test any combination of equality and inequality constraints on the variances. It is straightforward to simultaneously test multiple hypotheses. The aFBF then indicates which hypothesis receives strongest support from the data. In doing so it functions as Occam's razor by taking the parsimony of (in)equality constrained hypotheses into account. The aFBF is fully automatic, which means that the user does not need to specify a prior distribution under every hypothesis to be tested. The results of the simulation study indicate that the aFBF is consistent in the sense that it selects the true hypothesis as the sample size increases. This also holds for instances in which the true order of the population variances is slightly different from the hypothesized order. In this case the aFBF chooses the complement over the order constrained hypothesis as the sample size increases. The aFBF can be computed easily and quickly using our Shiny application.

In the multiplication and the division domain of the Math Garden the variances increased monotonically across grades as suggested by Aunola et al. (2004). In the addition and the subtraction domain, however, the variances first decreased from Grade 1 to Grade 2, followed by an increase over the years. Interestingly, both patterns are in line with a random slope model of development over time. Our approach can be used to test these and other variance patterns implied by models of development over time such as random slope and random quadratic models using cross-sectional data.

Like many other statistical methods, the aFBF assumes that the data are normally distributed. However, the normal distribution may not be an appropriate model for data that contain outliers or depart in other ways from normality (e.g., skewness and/or kurtosis). The robustness of the aFBF to such violations of normality is an

important topic for future study. Furthermore, it would be interesting to investigate how the aFBF behaves under conditions that differ from those in our simulation study. For example, in the simulation we assumed that the ratio of adjacent population variances is constant. Real-life psychological phenomena may involve more complex variance patterns, which is why further research investigating the behavior of the aFBF under different population variance structures is indicated.

In this chapter we focused on testing variances of independent groups. It appears natural to also consider the Bayes factor for testing variances of dependent groups since these are frequently encountered by psychologists. Such a method would be useful for analyzing repeated measurement data and other types of data where there is a relationship between the respondents of different groups. Our approach can be extended to dependent observations using a multivariate normal model $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the covariance matrix of the dependent measures. Constrained hypotheses are then formulated on the diagonal elements of this covariance matrix. The additional challenge in the dependent case is that the constraints on the variances are added to the constraints that ensure that the covariance matrix is positive definite. This is an interesting topic for future research.

3.A Fractional Bayes Factor for an Inequality Constrained Hypothesis Test

We consider the test of an inequality constrained hypothesis H_t on the variances of J groups against the unconstrained hypothesis $H_u\colon \sigma_1^2, \ldots, \sigma_J^2$. The inequality constrained hypothesis can be formulated as $H_t\colon \mathbf{R}_t\boldsymbol{\sigma}^2 > \mathbf{0}$, where the rows of $\mathbf{R}_t$ are permutations of $(1, -1, 0, \ldots, 0)$. For example, under $H_1\colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ the matrix is given by $\mathbf{R}_1 = \left[\begin{smallmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{smallmatrix}\right]$. The admissible parameter space of the group variances under H_t and H_u can be written as $\Omega_t = \left\{\boldsymbol{\sigma}^2 | \mathbf{R}_t\boldsymbol{\sigma}^2 > \mathbf{0}\right\}$ and $\Omega_u = (\mathbb{R}^+)^J$, respectively. Note that the likelihood and the noninformative improper prior under H_t are truncations of the unconstrained likelihood and prior:

$$f_t\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) = f_u\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) I_{\Omega_t}\left(\boldsymbol{\sigma}^2\right), \tag{3.30}$$

$$f_t\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^b = f_u\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^b I_{\Omega_t}\left(\boldsymbol{\sigma}^2\right), \text{ and} \tag{3.31}$$

$$\pi_t^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) = C_t\, \pi_u^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) I_{\Omega_t}\left(\boldsymbol{\sigma}^2\right), \tag{3.32}$$

where $I_{\Omega_t}\left(\boldsymbol{\sigma}^2\right)$ is an indicator function that equals 1 if $\boldsymbol{\sigma}^2 \in \Omega_t$ and 0 otherwise, and C_t is a normalizing constant. The FBF for an inequality constrained hypothesis H_t

against the unconstrained hypothesis H_u can then be written as

$$
\begin{aligned}
B_{tu}^F &= \frac{m_t^F(\mathbf{x}, b)}{m_u^F(\mathbf{x}, b)} \\
&= \frac{\frac{\int_{\Omega_t}\int_{\mathbb{R}^J} f_t\left(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)\pi_t^N\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)d\boldsymbol{\mu}\,d\boldsymbol{\sigma}^2}{\int_{\Omega_t}\int_{\mathbb{R}^J} f_t(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2)^b\pi_t^N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)d\boldsymbol{\mu}\,d\boldsymbol{\sigma}^2}}{\frac{\int_{\Omega_u}\int_{\mathbb{R}^J} f_u(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2)\pi_u^N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)d\boldsymbol{\mu}\,d\boldsymbol{\sigma}^2}{\int_{\Omega_u}\int_{\mathbb{R}^J} f_u(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2)^b\pi_u^N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)d\boldsymbol{\mu}\,d\boldsymbol{\sigma}^2}} = \frac{\frac{\int_{\Omega_t}\int_{\mathbb{R}^J} f_u\left(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)C_t\pi_u^N\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)I_{\Omega_t}\left(\boldsymbol{\sigma}^2\right)d\boldsymbol{\mu}\,d\boldsymbol{\sigma}^2}{\int_{\Omega_u}\int_{\mathbb{R}^J} f_u(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2)\pi_u^N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)d\boldsymbol{\mu}\,d\boldsymbol{\sigma}^2}}{\frac{\int_{\Omega_t}\int_{\mathbb{R}^J} f_u(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2)^bC_t\pi_u^N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)I_{\Omega_t}(\boldsymbol{\sigma}^2)d\boldsymbol{\mu}\,d\boldsymbol{\sigma}^2}{\int_{\Omega_u}\int_{\mathbb{R}^J} f_u(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2)^b\pi_u^N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)d\boldsymbol{\mu}\,d\boldsymbol{\sigma}^2}} \\
&= \frac{\int_{\Omega_t}\int_{\mathbb{R}^J}\frac{f_u\left(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)\pi_u^N\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)}{\int_{\Omega_u}\int_{\mathbb{R}^J} f_u(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2)\pi_u^N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)d\boldsymbol{\mu}\,d\boldsymbol{\sigma}^2}\,d\boldsymbol{\mu}\,d\boldsymbol{\sigma}^2}{\int_{\Omega_t}\int_{\mathbb{R}^J}\frac{f_u(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2)^b\pi_u^N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)}{\int_{\Omega_u}\int_{\mathbb{R}^J} f_u(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2)^b\pi_u^N(\boldsymbol{\mu},\boldsymbol{\sigma}^2)d\boldsymbol{\mu}\,d\boldsymbol{\sigma}^2}\,d\boldsymbol{\mu}\,d\boldsymbol{\sigma}^2} = \frac{\int_{\Omega_t}\int_{\mathbb{R}^J}\pi_u\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2|\mathbf{x}\right)d\boldsymbol{\mu}\,d\boldsymbol{\sigma}^2}{\int_{\Omega_t}\int_{\mathbb{R}^J}\pi_u\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2|\mathbf{x}^b\right)d\boldsymbol{\mu}\,d\boldsymbol{\sigma}^2} \\
&= \frac{\int_{\Omega_t}\pi_u\left(\boldsymbol{\sigma}^2|\mathbf{x}\right)d\boldsymbol{\sigma}^2}{\int_{\Omega_t}\pi_u\left(\boldsymbol{\sigma}^2|\mathbf{x}^b\right)d\boldsymbol{\sigma}^2} = \frac{P\left(\boldsymbol{\sigma}^2\in\Omega_t|\mathbf{x}\right)}{P\left(\boldsymbol{\sigma}^2\in\Omega_t|\mathbf{x}^b\right)}.
\end{aligned}
\tag{3.33}
$$

Note that in the second line the indicator function $I_{\Omega_t}\left(\boldsymbol{\sigma}^2\right)$ in the constrained likelihood and prior, f_t and π_t^N, respectively, can be omitted because the integration region is already restricted to the constrained parameter space Ω_t.

3.B Computation of the Marginal Likelihood in the Adjusted Fractional Bayes Factor

We consider a hypothesis H_t with equality and inequality constraints on the variances. We have to introduce some additional notation before deriving the marginal likelihood in the aFBF approach. Under H_t, let there be q^E equality constraints and q^I inequality constraints on the variances, where we omitted the hypothesis index t on q^E and q^I to simplify the notation. Thus, there are $K = J - q^E$ unique variances under H_t. We denote these K unique variances by $\tilde{\boldsymbol{\sigma}}^2 = \left(\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_K^2\right)'$. The q^I inequality constraints are formulated on these unique variances. Furthermore let J_k be the number of groups that share the unique variance $\tilde{\sigma}_k^2$, and let $\mathbf{x}_{k_j}$, μ_{k_j}, and n_{k_j} denote the data, the mean, and the sample size of the jth group sharing the unique variance $\tilde{\sigma}_k^2$, respectively.

For example, consider the hypothesis $H_1\colon \sigma_1^2 = \sigma_2^2 < \sigma_3^2 = \sigma_4^2$ on the variances of $J = 4$ groups. Under H_1 there are $q^E = 2$ equality constraints and $q^I = 1$ inequality constraint, so that the number of unique variances is given by $K = 4 - 2 = 2$. We denote these variances by $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$. Then Groups 1 and 2 have unique variance $\tilde{\sigma}_1^2$ and Groups 3 and 4 have unique variance $\tilde{\sigma}_2^2$. Thus, hypothesis H_1 can be written as $H_1\colon \tilde{\sigma}_1^2 < \tilde{\sigma}_2^2$. Furthermore, we have $(J_1, J_2)' = (2, 2)'$. In this notation, $\mathbf{x}_{1_1}$, $\mathbf{x}_{1_2}$, $\mathbf{x}_{2_1}$, and $\mathbf{x}_{2_2}$ correspond to the data of Group 1, 2, 3, and 4, respectively, μ_{1_1}, μ_{1_2}, μ_{2_1}, and μ_{2_2} are the means of Group 1, 2, 3, and 4, and n_{1_1}, n_{1_2}, n_{2_1}, and n_{2_2} are the sample sizes of Group 1, 2, 3, and 4.

The marginal likelihood under a constrained hypothesis H_t in the adjusted frac-

tional Bayes factor is defined by

$$m_t^{aF}(\mathbf{x},\mathbf{b}) = \frac{\int_{\Omega_t}\int_{\mathbb{R}^J} f_u\left(\mathbf{x}|\boldsymbol{\mu},\tilde{\boldsymbol{\sigma}}^2\right)\pi_u^N\left(\boldsymbol{\mu},\tilde{\boldsymbol{\sigma}}^2\right)d\boldsymbol{\mu}\,d\tilde{\boldsymbol{\sigma}}^2}{\int_{\Omega_t^a}\int_{\mathbb{R}^J} f_u\left(\mathbf{x}|\boldsymbol{\mu},\tilde{\boldsymbol{\sigma}}^2\right)^{\mathbf{b}}\pi_u^N\left(\boldsymbol{\mu},\tilde{\boldsymbol{\sigma}}^2\right)d\boldsymbol{\mu}\,d\tilde{\boldsymbol{\sigma}}^2} = \frac{m_t^N(\mathbf{x})}{m_t^N(\mathbf{x}^{\mathbf{b}})}, \tag{3.34}$$

where the likelihood and the noninformative prior are used without the inequality constraints on the unique variances $\tilde{\boldsymbol{\sigma}}^2$, which is part of the definition of the aFBF. The expressions are given by

$$f_u\left(\mathbf{x}|\boldsymbol{\mu},\tilde{\boldsymbol{\sigma}}^2\right) = \prod_{k=1}^{K}\prod_{j=1}^{J_k} f\left(\mathbf{x}_{k_j}|\mu_{k_j},\tilde{\sigma}_k^2\right), \tag{3.35}$$

$$f_u\left(\mathbf{x}|\boldsymbol{\mu},\tilde{\boldsymbol{\sigma}}^2\right)^{\mathbf{b}} = \prod_{k=1}^{K}\prod_{j=1}^{J_k} f\left(\mathbf{x}_{k_j}|\mu_{k_j},\tilde{\sigma}_k^2\right)^{b_{k_j}}, \text{ and} \tag{3.36}$$

$$\pi_u^N\left(\boldsymbol{\mu},\tilde{\boldsymbol{\sigma}}^2\right) = C_t\prod_{k=1}^{K}\tilde{\sigma}_k^{-2}, \tag{3.37}$$

with

$$f\left(\mathbf{x}_{k_j}|\mu_{k_j},\tilde{\sigma}_k^2\right) = \left(\tilde{\sigma}_k^2 2\pi\right)^{-\frac{n_{k_j}}{2}}\exp\left(-\frac{1}{2\tilde{\sigma}_k^2}\left((n_{k_j}-1)s_{k_j}^2 + n_{k_j}(\bar{x}_{k_j}-\mu_{k_j})^2\right)\right) \text{ and} \tag{3.38}$$

$$f\left(\mathbf{x}_{k_j}|\mu_{k_j},\tilde{\sigma}_k^2\right)^{b_{k_j}} = \left(\tilde{\sigma}_k^2 2\pi\right)^{-\frac{b_{k_j}n_{k_j}}{2}}\exp\left(-\frac{b_{k_j}}{2\tilde{\sigma}_k^2}\left((n_{k_j}-1)s_{k_j}^2 + n_{k_j}(\bar{x}_{k_j}-\mu_{k_j})^2\right)\right). \tag{3.39}$$

Note that Equation (3.34) is identical to Equation (3.15) except that a tilde is used for the unique variances which are integrated out.

The constrained parameter space Ω_t in the numerator in Equation (3.34) can be written as

$$\Omega_t = \left\{\tilde{\boldsymbol{\sigma}}^2 \,|\, \mathbf{R}_t\left(\tilde{\sigma}_1^2,\ldots,\tilde{\sigma}_K^2\right)' > \mathbf{0}\right\}, \tag{3.40}$$

where the rows of $\mathbf{R}_t$ are permutations of $(1,-1,0,\ldots,0)$. For example, under $H_1\colon \tilde{\sigma}_1^2 < \tilde{\sigma}_2^2$ the matrix is given by $\mathbf{R}_1 = \begin{bmatrix}-1 & 1\end{bmatrix}$. The adjusted constrained parameter space Ω_t^a in the denominator in Equation (3.34), which is a crucial part of the aFBF approach, can be written as

$$\Omega_t^a = \left\{\tilde{\boldsymbol{\sigma}}^2 \,|\, \mathbf{R}_t\left(a_1\tilde{\sigma}_1^2,\ldots,a_K\tilde{\sigma}_K^2\right)' > \mathbf{0}\right\}, \tag{3.41}$$

where the tuning parameters are set to

$$a_k = \frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}{\sum_{j=1}^{J_k} b_{k_j} \left(n_{k_j} - 1\right) s_{k_j}^2}, \tag{3.42}$$

for $k = 1, \ldots, K$. This tuning results in equal scale hyperparameters in the automatic prior for the unique variances. This will be shown after the derivation of the marginal likelihood.

We first derive the denominator $m_t^N(\mathbf{x}^{\mathbf{b}})$ of the marginal likelihood in Equation (3.34). Substituting the expressions for the fraction of the likelihood and the Jeffreys prior in Equations (3.36) and (3.37) into the denominator of Equation (3.34) gives us

$$\begin{aligned}
&m_t^N(\mathbf{x}^{\mathbf{b}}) \\
&\quad = \int_{\Omega_t^a} \int_{\mathbb{R}^J} \left(\prod_{k=1}^{K} \prod_{j=1}^{J_k} f\left(\mathbf{x}_{k_j} | \mu_{k_j}, \tilde{\sigma}_k^2\right)^{b_{k_j}} \right) C_t \prod_{k=1}^{K} \tilde{\sigma}_k^{-2} \, d\boldsymbol{\mu} \, d\tilde{\boldsymbol{\sigma}}^2 \\
&\quad = C_t \int_{\Omega_t^a} \prod_{k=1}^{K} \tilde{\sigma}_k^{-2} \prod_{j=1}^{J_k} \int_{\mathbb{R}} \left(\tilde{\sigma}_k^2 2\pi\right)^{-\frac{b_{k_j} n_{k_j}}{2}} \\
&\qquad\qquad \exp\left(-\frac{b_{k_j}}{2\tilde{\sigma}_k^2} \left(\left(n_{k_j} - 1\right) s_{k_j}^2 + n_{k_j} \left(\bar{x}_{k_j} - \mu_{k_j}\right)^2 \right) \right) d\mu_{k_j} \, d\tilde{\boldsymbol{\sigma}}^2 \\
&\quad = C_t \int_{\Omega_t^a} \prod_{k=1}^{K} \tilde{\sigma}_k^{-2} \prod_{j=1}^{J_k} \left(b_{k_j} n_{k_j}\right)^{-\frac{1}{2}} \left(\tilde{\sigma}_k^2 2\pi\right)^{-\frac{b_{k_j} n_{k_j} - 1}{2}} \exp\left(-\frac{b_{k_j} \left(n_{k_j} - 1\right) s_{k_j}^2}{2\tilde{\sigma}_k^2} \right) d\tilde{\boldsymbol{\sigma}}^2 \\
&\quad = C_t \left(\prod_{k=1}^{K} \prod_{j=1}^{J_k} \left(b_{k_j} n_{k_j}\right)^{-\frac{1}{2}} \right) (2\pi)^{-\frac{\sum_{k=1}^{K} \left(\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k \right)}{2}} \\
&\qquad\qquad \int_{\Omega_t^a} \prod_{k=1}^{K} \left(\tilde{\sigma}_k^2\right)^{-\left(\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}{2} + 1 \right)} \exp\left(-\frac{\sum_{j=1}^{J_k} b_{k_j} \left(n_{k_j} - 1\right) s_{k_j}^2}{2\tilde{\sigma}_k^2} \right) d\tilde{\boldsymbol{\sigma}}^2 \\
&\quad = C_t \left(\prod_{k=1}^{K} \prod_{j=1}^{J_k} \left(b_{k_j} n_{k_j}\right)^{-\frac{1}{2}} \right) \pi^{-\frac{\sum_{k=1}^{K} \left(\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k \right)}{2}} \\
&\qquad \left(\prod_{k=1}^{K} \Gamma\left(\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}{2} \right) \left(\sum_{j=1}^{J_k} b_{k_j} \left(n_{k_j} - 1\right) s_{k_j}^2 \right)^{-\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}{2}} \right) \\
&\qquad\qquad \int_{\Omega_t^a} \prod_{k=1}^{K} \text{Inv-}\chi^2\left(\tilde{\sigma}_k^2 \middle| \left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j} \right) - J_k, \frac{\sum_{j=1}^{J_k} b_{k_j} \left(n_{k_j} - 1\right) s_{k_j}^2}{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k} \right) d\tilde{\boldsymbol{\sigma}}^2
\end{aligned}$$

$$= C_t \left(\prod_{k=1}^{K} \prod_{j=1}^{J_k} \left(b_{k_j} n_{k_j} \right)^{-\frac{1}{2}} \right) \pi^{-\frac{\sum_{k=1}^{K} \left(\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j} \right) - J_k \right)}{2}} \left(\prod_{k=1}^{K} \Gamma \left(\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j} \right) - J_k}{2} \right) \left(\sum_{j=1}^{J_k} b_{k_j} \left(n_{k_j} - 1 \right) s_{k_j}^2 \right)^{-\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j} \right) - J_k}{2}} \right) P \left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a | \mathbf{x}^{\mathbf{b}} \right). \tag{3.43}$$

In the second line we solved the integral with respect to μ_{k_j} by integrating $\exp\left(-\frac{b_{k_j} n_{k_j}}{2\tilde{\sigma}_k^2} \left(\mu_{k_j} - \bar{x}_{k_j} \right)^2 \right)$, which is the kernel of a normal distribution with mean $\bar{x}_{k_j}$ and variance $\tilde{\sigma}_k^2 / \left(b_{k_j} n_{k_j} \right)$. Hence the integral equals $\left(2\pi \tilde{\sigma}_k^2 / \left(b_{k_j} n_{k_j} \right) \right)^{\frac{1}{2}}$, which is the inverse of the normalizing constant of this normal distribution. The integrand in the fourth line is a product of kernels of scaled inverse-χ^2 distributions with degrees of freedom parameters $\nu_k = \left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j} \right) - J_k$ and scale parameters $\tau_k^2 = \frac{\sum_{j=1}^{J_k} b_{k_j} \left(n_{k_j} - 1 \right) s_{k_j}^2}{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j} \right) - J_k}$, $k = 1, \ldots, K$ (Gelman et al., 2004). Finally, the probability that the variances fall in the adjusted parameter space Ω_t^a is based on independent automatic priors for the variances given by

$$\tilde{\sigma}_k^2 | \mathbf{x}^{\mathbf{b}} \sim \text{Inv-}\chi^2 \left(\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j} \right) - J_k, \frac{\sum_{j=1}^{J_k} b_{k_j} \left(n_{k_j} - 1 \right) s_{k_j}^2}{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j} \right) - J_k} \right), \tag{3.44}$$

for $k = 1, \ldots, K$.

The expression for the numerator $m_t^N(\mathbf{x})$ of the marginal likelihood in Equation (3.34) is identical to the final expression in Equation (3.43) with all b's set to 1 and Ω_t^a replaced by Ω_t:

$$m_t^N(\mathbf{x}) = C_t \left(\prod_{k=1}^{K} \prod_{j=1}^{J_k} n_{k_j}^{-\frac{1}{2}} \right) \pi^{-\frac{\sum_{k=1}^{K} \left(\left(\sum_{j=1}^{J_k} n_{k_j} \right) - J_k \right)}{2}} \left(\prod_{k=1}^{K} \Gamma \left(\frac{\left(\sum_{j=1}^{J_k} n_{k_j} \right) - J_k}{2} \right) \left(\sum_{j=1}^{J_k} \left(n_{k_j} - 1 \right) s_{k_j}^2 \right)^{-\frac{\left(\sum_{j=1}^{J_k} n_{k_j} \right) - J_k}{2}} \right) P \left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t | \mathbf{x} \right). \tag{3.45}$$

Subsequently, the marginal likelihood in the aFBF is given by

$$m_t^{aF}(\mathbf{x}, \mathbf{b}) = \frac{m_t^N(\mathbf{x})}{m_t^N(\mathbf{x}^{\mathbf{b}})} = \tilde{m}_t^{aF}(\mathbf{x}, \mathbf{b}) \frac{P\left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t | \mathbf{x} \right)}{P\left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a | \mathbf{x}^{\mathbf{b}} \right)}, \tag{3.46}$$

where

$$\tilde{m}_t^{aF}(\mathbf{x},\mathbf{b}) = \left(\prod_{k=1}^{K}\prod_{j=1}^{J_k} b_{k_j}^{\frac{1}{2}}\right)\pi^{-\frac{\sum_{k=1}^{K}\sum_{j=1}^{J_k}\left(1-b_{k_j}\right)n_{k_j}}{2}} \prod_{k=1}^{K}\Gamma\left(\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}\right)\Gamma\left(\frac{\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right)-J_k}{2}\right)^{-1} \left(\sum_{j=1}^{J_k}\left(n_{k_j}-1\right)s_{k_j}^2\right)^{-\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}} \left(\sum_{j=1}^{J_k} b_{k_j}\left(n_{k_j}-1\right)s_{k_j}^2\right)^{\frac{\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right)-J_k}{2}}. \tag{3.47}$$

Note that if a constrained hypothesis does not contain any inequalities, the ratio of probabilities in Equation (3.46) is not present.

Finally, we provide a motivation for the specific choice of the tuning parameters. First we introduce new parameters $\phi_k = a_k\tilde{\sigma}_k^2$, for $k = 1,\ldots,K$, which can be interpreted as adjusted variance parameters. In the automatic prior the adjusted variance is distributed according to

$$\begin{aligned}\phi_k = a_k\tilde{\sigma}_k^2|\mathbf{x}^{\mathbf{b}} &\sim \text{Inv-}\chi^2\left(\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right)-J_k, a_k\frac{\sum_{j=1}^{J_k} b_{k_j}\left(n_{k_j}-1\right)s_{k_j}^2}{\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right)-J_k}\right)\\ &= \text{Inv-}\chi^2\left(\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right)-J_k, 1\right),\end{aligned} \tag{3.48}$$

for $k = 1,\ldots,K$, which follows automatically from Equations (3.42) and (3.44). In the first line we used the mathematical result that if $\tilde{\sigma}_k^2|\mathbf{x}^{\mathbf{b}} \sim \text{Inv-}\chi^2\left(\nu_k,\tau_k^2\right)$, then $a_k\tilde{\sigma}_k^2|\mathbf{x}^{\mathbf{b}} \sim \text{Inv-}\chi^2\left(\nu_k, a_k\tau_k^2\right)$. Note that the scale hyperparameters of the scaled inverse-χ^2 distributions are equal for all k. Subsequently, the automatic prior probability that the variances fall in the adjusted constrained space Ω_t^a can be written as

$$P\left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a|\mathbf{x}^{\mathbf{b}}\right) = P\left(\mathbf{R}_t\left(a_1\tilde{\sigma}_1^2,\ldots,a_K\tilde{\sigma}_K^2\right)' > \mathbf{0}|\mathbf{x}^{\mathbf{b}}\right) = P\left(\mathbf{R}_t(\phi_1,\ldots,\phi_K)' > \mathbf{0}|\mathbf{x}^{\mathbf{b}}\right). \tag{3.49}$$

To illustrate the effect of the adjustment we again consider the hypothesis $H_1\colon \tilde{\sigma}_1^2 < \tilde{\sigma}_2^2$ with $\mathbf{R}_1 = \begin{bmatrix}-1 & 1\end{bmatrix}$. If we set $b_{k_j} = 2/n_{k_j}$, the automatic prior probability that the variances fall in the adjusted constrained space equals

$$P\left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a|\mathbf{x}^{\mathbf{b}}\right) = P\left(a_1\tilde{\sigma}_1^2 < a_2\tilde{\sigma}_2^2|\mathbf{x}^{\mathbf{b}}\right) = P\left(\phi_1 < \phi_2|\mathbf{x}^{\mathbf{b}}\right) = \frac{1}{2}, \tag{3.50}$$

because ϕ_1 and ϕ_2 are both distributed as Inv-$\chi^2(2,1)$ due to Equation (3.48). This is desirable because it implies that in the aFBF approach both possible orderings of the two adjusted variances are equally likely a priori.

3.C Scale Invariance of the Adjusted Fractional Bayes Factor

In Appendix 3.B the data and the sample variance of the jth group sharing the unique variance $\tilde{\sigma}_k^2$ were denoted by $\mathbf{x}_{k_j}$ and $s_{k_j}^2$, respectively. Multiplying all observations in $\mathbf{x}_{k_j}$ by a constant w results in a sample variance of $w^2 s_{k_j}^2$, for $j = 1, \ldots, J_k$ and $k = 1, \ldots, K$. We show that the marginal likelihood of the scaled data $w\mathbf{x}$ under hypothesis H_t can be written as $m_t^{aF}(w\mathbf{x}, \mathbf{b}) = v\, m_t^{aF}(\mathbf{x}, \mathbf{b})$, where v is a constant that is independent of H_t. As will be shown, v cancels out in the computation of the adjusted fractional Bayes factors and the corresponding posterior probabilities of the hypotheses. We first consider $\tilde{m}_t^{aF}(w\mathbf{x}, \mathbf{b})$ in Equation (3.46) for the scaled data. Note that the marginal likelihood only depends on the data through the sample variances. Thus, substituting $s_{k_j}^2$ with $w^2 s_{k_j}^2$ in Equation (3.47) gives us

$$
\begin{aligned}
&\tilde{m}_t^{aF}(w\mathbf{x}, \mathbf{b}) \\
&= \left(\prod_{k=1}^{K} \prod_{j=1}^{J_k} b_{k_j}^{\frac{1}{2}} \right) \pi^{-\frac{\sum_{k=1}^{K} \sum_{j=1}^{J_k} \left(1 - b_{k_j}\right) n_{k_j}}{2}} \\
&\quad \prod_{k=1}^{K} \Gamma\left(\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}{2} \right) \Gamma\left(\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}{2} \right)^{-1} \\
&\quad \left(\sum_{j=1}^{J_k} \left(n_{k_j} - 1\right) w^2 s_{k_j}^2 \right)^{-\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}{2}} \left(\sum_{j=1}^{J_k} b_{k_j} \left(n_{k_j} - 1\right) w^2 s_{k_j}^2 \right)^{\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}{2}} \\
&= \left(w^2\right)^{-\frac{\sum_{k=1}^{K} \sum_{j=1}^{J_k} \left(1 - b_{k_j}\right) n_{k_j}}{2}} \left(\prod_{k=1}^{K} \prod_{j=1}^{J_k} b_{k_j}^{\frac{1}{2}} \right) \pi^{-\frac{\sum_{k=1}^{K} \sum_{j=1}^{J_k} \left(1 - b_{k_j}\right) n_{k_j}}{2}} \\
&\quad \prod_{k=1}^{K} \Gamma\left(\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}{2} \right) \Gamma\left(\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}{2} \right)^{-1} \\
&\quad \left(\sum_{j=1}^{J_k} \left(n_{k_j} - 1\right) s_{k_j}^2 \right)^{-\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}{2}} \left(\sum_{j=1}^{J_k} b_{k_j} \left(n_{k_j} - 1\right) s_{k_j}^2 \right)^{\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}{2}} \\
&= \left(w^2\right)^{-\frac{\sum_{k=1}^{K} \sum_{j=1}^{J_k} \left(1 - b_{k_j}\right) n_{k_j}}{2}} \tilde{m}_t^{aF}(\mathbf{x}, \mathbf{b}) = v\, \tilde{m}_t^{aF}(\mathbf{x}, \mathbf{b}).
\end{aligned}
\tag{3.51}
$$

Next, we consider $P\left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t | w\mathbf{x}\right)$ and $P\left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a | (w\mathbf{x})^{\mathbf{b}}\right)$ in Equation (3.46) for the scaled data. For the scaled data the variances are distributed according to

$$
\tilde{\sigma}_k^2 | (w\mathbf{x})^{\mathbf{b}} \sim \text{Inv-}\chi^2 \left(\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j} \right) - J_k, w^2 \frac{\sum_{j=1}^{J_k} b_{k_j} \left(n_{k_j} - 1\right) s_{k_j}^2}{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k} \right) \tag{3.52}
$$

and

$$\tilde{\sigma}_k^2|w\mathbf{x} \sim \text{Inv-}\chi^2\left(\left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k, w^2\frac{\sum_{j=1}^{J_k}\left(n_{k_j}-1\right)s_{k_j}^2}{\left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}\right), \tag{3.53}$$

for the automatic prior and posterior, respectively, for $k = 1, \ldots, K$. Because the scale parameters in the above distributions only depend on the scale w through the factor w^2, it automatically follows that the automatic prior probability is invariant of the scale, that is,

$$\begin{aligned} P\left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a|(w\mathbf{x})^{\mathbf{b}}\right) &= P\left(w^2\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a|\mathbf{x}^{\mathbf{b}}\right) = P\left(\mathbf{R}_t\left(w^2a_1\tilde{\sigma}_1^2, \ldots, w^2a_K\tilde{\sigma}_K^2\right)' > \mathbf{0}|\mathbf{x}^{\mathbf{b}}\right) \\ &= P\left(\mathbf{R}_t\left(a_1\tilde{\sigma}_1^2, \ldots, a_K\tilde{\sigma}_K^2\right)' > \mathbf{0}|\mathbf{x}^{\mathbf{b}}\right) = P\left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a|\mathbf{x}^{\mathbf{b}}\right). \end{aligned} \tag{3.54}$$

By following the same steps it can be shown that the posterior probability is also invariant, that is, $P\left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t|w\mathbf{x}\right) = P\left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t|\mathbf{x}\right)$. Thus, the marginal likelihood of the scaled data can be written as

$$\begin{aligned} m_t^{aF}(w\mathbf{x}, \mathbf{b}) &= \tilde{m}_t^{aF}(w\mathbf{x}, \mathbf{b})\frac{P\left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t|w\mathbf{x}\right)}{P\left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a|(w\mathbf{x})^{\mathbf{b}}\right)} \\ &= v\,\tilde{m}_t^{aF}(\mathbf{x}, \mathbf{b})\frac{P\left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t|\mathbf{x}\right)}{P\left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a|\mathbf{x}^{\mathbf{b}}\right)} = v\,m_t^{aF}(\mathbf{x}, \mathbf{b}). \end{aligned} \tag{3.55}$$

Since the constant v is the same under all hypotheses, it cancels out in the computation of the adjusted fractional Bayes factors and the corresponding posterior probabilities of the hypotheses:

$$B_{tt'}^{aF} = \frac{m_t^{aF}(w\mathbf{x}, \mathbf{b})}{m_{t'}^{aF}(w\mathbf{x}, \mathbf{b})} = \frac{v\,m_t^{aF}(\mathbf{x}, \mathbf{b})}{v\,m_{t'}^{aF}(\mathbf{x}, \mathbf{b})} = \frac{m_t^{aF}(\mathbf{x}, \mathbf{b})}{m_{t'}^{aF}(\mathbf{x}, \mathbf{b})} \tag{3.56}$$

and

$$\begin{aligned} P^{aF}(H_t|w\mathbf{x}, \mathbf{b}) &= \frac{m_t^{aF}(w\mathbf{x}, \mathbf{b})P(H_t)}{\sum_{t'=1}^{T} m_{t'}^{aF}(w\mathbf{x}, \mathbf{b})P(H_{t'})} = \frac{v\,m_t^{aF}(\mathbf{x}, \mathbf{b})P(H_t)}{\sum_{t'=1}^{T} v\,m_t^{aF}(\mathbf{x}, \mathbf{b})P(H_{t'})} \\ &= \frac{m_t^{aF}(\mathbf{x}, \mathbf{b})P(H_t)}{\sum_{t'=1}^{T} m_{t'}^{aF}(\mathbf{x}, \mathbf{b})P(H_{t'})} = P^{aF}(H_t|\mathbf{x}, \mathbf{b}). \end{aligned} \tag{3.57}$$

3.D Supplemental Material

Figures 3.10 to 3.14 show the simulation results for the unequal group sizes given in Table 3.7; all other settings were identical to the simulation with equal group sizes. The relative group size is constant across conditions: For $J = 3$ groups relative group sizes are given by $3/15 = 0.20$, $5/15 = 0.33$, and $7/15 = 0.47$, for group 1, 2, and 3, respectively. In the $J = 5$ groups case the relative group sizes are $3/25 = 0.12$, $4/25 = 0.16$, $5/25 = 0.20$, $6/25 = 0.24$, and $7/25 = 0.28$, for group 1, 2, 3, 4, and 5, respectively.

Table 3.7: Sample sizes in the simulation with an unbalanced design for $J = 3$ and 5 groups.

		Condition									
Number of groups	n_j	1	2	3	4	5	6	7	8	9	10
$J = 3$ groups	n_1	3	6	12	30	60	120	300	600	1200	3000
	n_2	5	10	20	50	100	200	500	1000	2000	5000
	n_3	7	14	28	70	140	280	700	1400	2800	7000
$J = 5$ groups	n_1	3	6	12	30	60	120	300	600	1200	3000
	n_2	4	8	16	40	80	160	400	800	1600	4000
	n_3	5	10	20	50	100	200	500	1000	2000	5000
	n_4	6	12	24	60	120	240	600	1200	2400	6000
	n_5	7	14	28	70	140	280	700	1400	2800	7000
Average sample size	n	5	10	20	50	100	200	500	1000	2000	5000

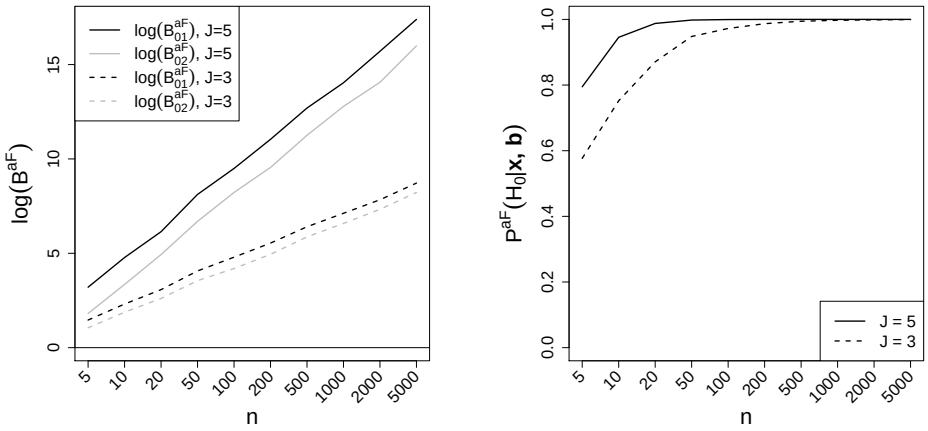


Figure 3.10: Simulation results for a null population in which all population variances were equal, $\sigma_1^2 = \cdots = \sigma_J^2$, for $J = 3$ groups (dashed lines) and $J = 5$ groups (solid lines). We tested the true hypothesis $H_0 \colon \sigma_1^2 = \cdots = \sigma_J^2$ against the two competing hypotheses $H_1 \colon \sigma_1^2 < \cdots < \sigma_J^2$ and $H_2 \colon$ not $\sigma_1^2 < \cdots < \sigma_J^2$. The plots show the median log Bayes factors (left-hand column) testing H_0 against H_1 (black lines) and H_0 against H_2 (gray lines) and the median posterior probability of H_0 (right-hand column) as a function of the average sample size n.

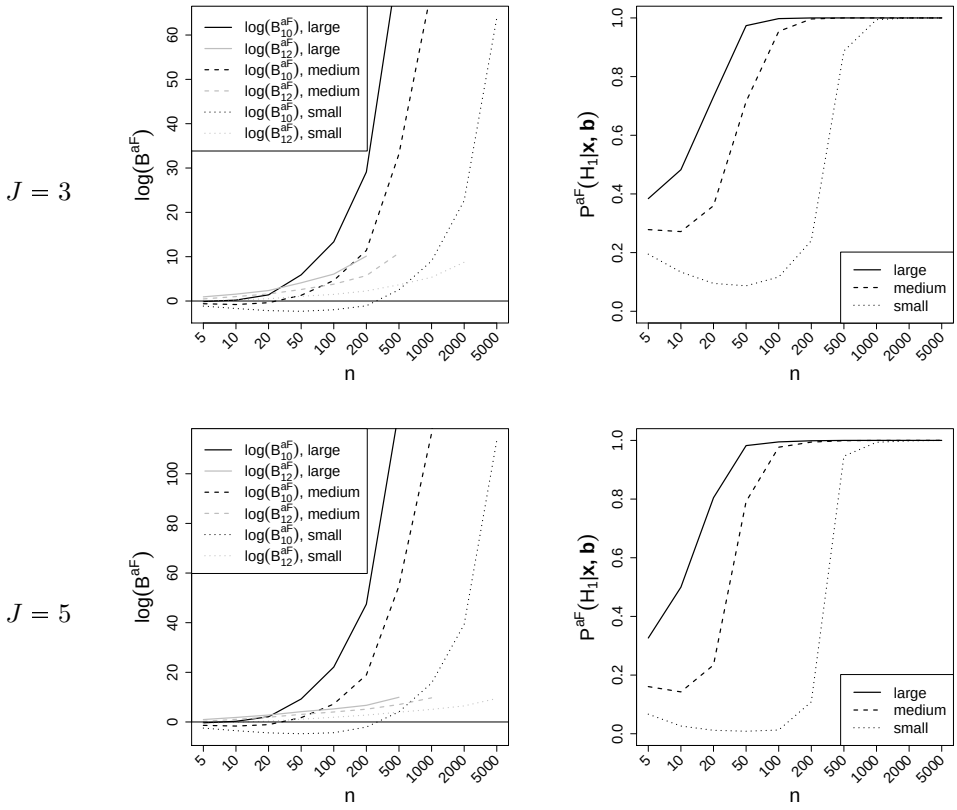


Figure 3.11: Simulation results for an order population in which the structure of the population variances was $\sigma_1^2 < \cdots < \sigma_J^2$, with $J \in \{3, 5\}$. We considered three effect sizes: small (dotted lines), medium (dashed lines), and large (solid lines). We tested the true hypothesis $H_1 \colon \sigma_1^2 < \cdots < \sigma_J^2$ against the two competing hypotheses $H_0 \colon \sigma_1^2 = \cdots = \sigma_J^2$ and $H_2 \colon$ not $\sigma_1^2 < \cdots < \sigma_J^2$. The plots show the median log Bayes factors (left-hand column) testing H_1 against H_0 (black lines) and H_1 against H_2 (gray lines) and the median posterior probability of H_1 (right-hand column) as a function of the average sample size n. In the log Bayes factors plots the gray lines are discontinued due to numerical reasons (see text).

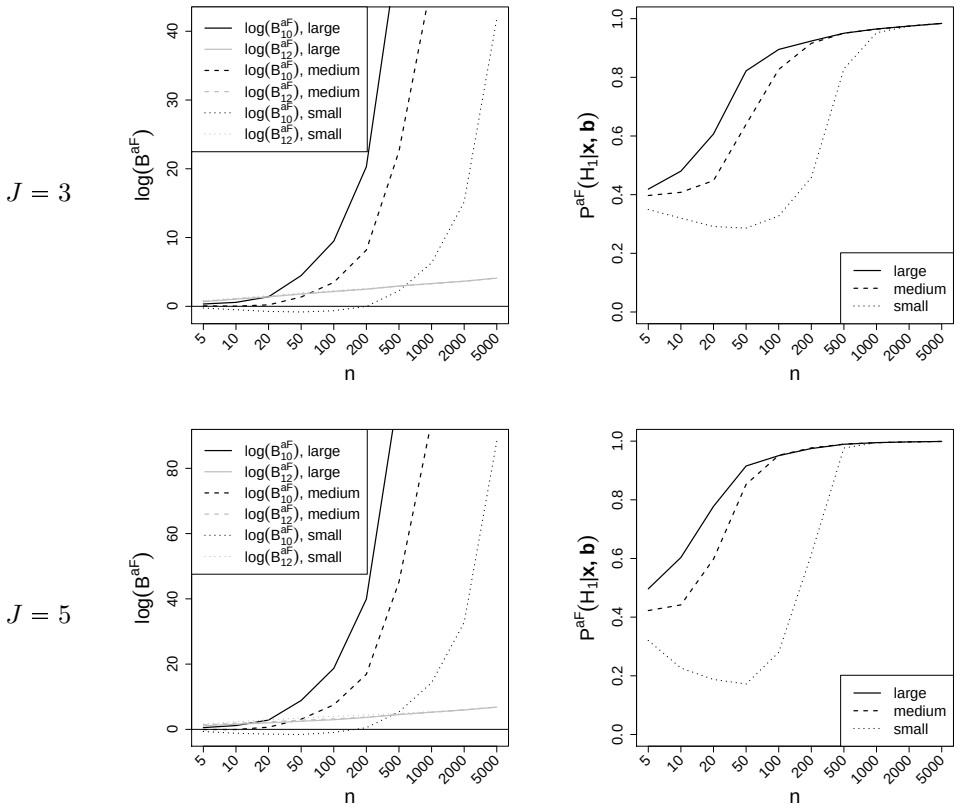


Figure 3.12: Simulation results for a mixed population. For $J = 3$ groups the structure of the population variances was $\sigma_1^2 < \sigma_2^2 = \sigma_3^2$, whereas for $J = 5$ groups it was $\sigma_1^2 < \sigma_2^2 = \sigma_3^2 < \sigma_4^2 = \sigma_5^2$. We considered three effect sizes: small (dotted lines), medium (dashed lines), and large (solid lines). For $J = 3$ groups we tested the true hypothesis $H_1 \colon \sigma_1^2 < \sigma_2^2 = \sigma_3^2$ against the two competing hypotheses $H_0 \colon \sigma_1^2 = \sigma_2^2 = \sigma_3^2$ and $H_2 \colon \sigma_1^2 < \left(\sigma_2^2, \sigma_3^2\right)$. For $J = 5$ groups we tested the true hypothesis $H_1 \colon \sigma_1^2 < \sigma_2^2 = \sigma_3^2 < \sigma_4^2 = \sigma_5^2$ against $H_0 \colon \sigma_1^2 = \cdots = \sigma_5^2$ and $H_2 \colon \sigma_1^2 < \left(\sigma_2^2, \sigma_3^2\right) < \left(\sigma_4^2, \sigma_5^2\right)$. The plots show the median log Bayes factors (left-hand column) testing H_1 against H_0 (black lines) and H_1 against H_2 (gray lines) and the median posterior probability of H_1 (right-hand column) as a function of the average sample size n.

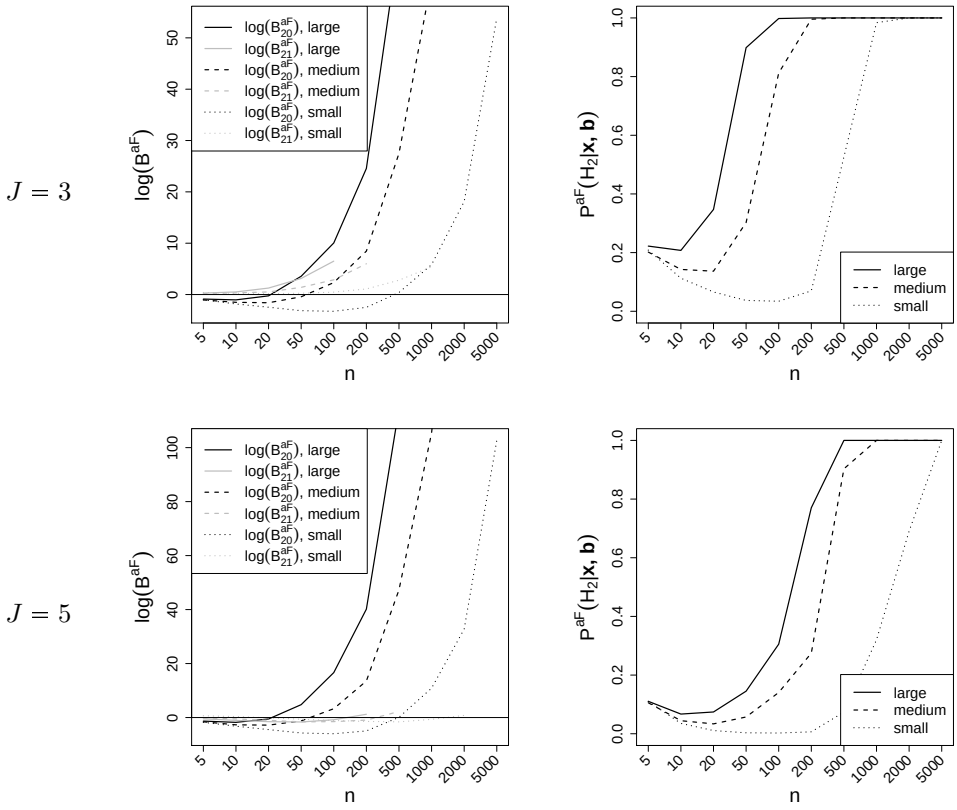


Figure 3.13: Simulation results for a near order population in which the structure of the population variances was $\sigma_1^2 < \cdots < \sigma_J^2 < \sigma_{J-1}^2$, with $J \in \{3, 5\}$. We considered three effect sizes: small (dotted lines), medium (dashed lines), and large (solid lines). We tested three hypotheses: $H_0\colon \sigma_1^2 = \cdots = \sigma_J^2$, $H_1\colon \sigma_1^2 < \cdots < \sigma_J^2$, and $H_2\colon$ not $\sigma_1^2 < \cdots < \sigma_J^2$. Note that the true hypothesis is the complement H_2. The plots show the median log Bayes factors (left-hand column) testing H_2 against H_0 (black lines) and H_2 against H_1 (gray lines) and the median posterior probability of H_2 (right-hand column) as a function of the average sample size n. In the log Bayes factors plots the gray lines are discontinued due to numerical reasons (see text).

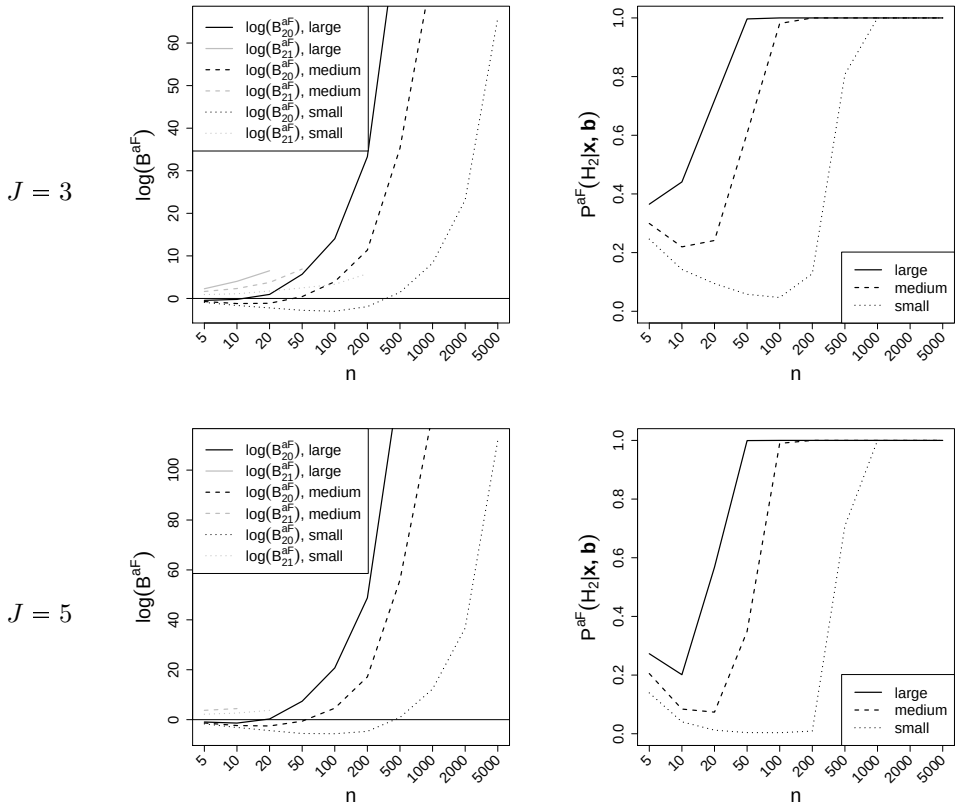


Figure 3.14: Simulation results for a reverse order population in which the structure of the population variances was $\sigma_J^2 < \cdots < \sigma_1^2$, with $J \in \{3, 5\}$. We considered three effect sizes: small (dotted lines), medium (dashed lines), and large (solid lines). We tested three hypotheses: H_0: $\sigma_1^2 = \cdots = \sigma_J^2$, H_1: $\sigma_1^2 < \cdots < \sigma_J^2$, and H_2: not $\sigma_1^2 < \cdots < \sigma_J^2$. Note that the true hypothesis is the complement H_2. The plots show the median log Bayes factors (left-hand column) testing H_2 against H_0 (black lines) and H_2 against H_1 (gray lines) and the median posterior probability of H_2 (right-hand column) as a function of the average sample size n. In the log Bayes factors plots the gray lines are discontinued due to numerical reasons (see text).

Chapter 4

Automatic Bayes Factors for Testing Equality and Inequality Constrained Hypotheses on Variances

Abstract

In comparing characteristics of independent populations, researchers frequently expect a certain structure of the population variances. These expectations can be formulated as hypotheses with equality and/or inequality constraints on the variances. In this chapter we consider the Bayes factor for testing such (in)equality constrained hypotheses on variances. Application of Bayes factors requires specification of a prior under every hypothesis to be tested. However, specifying subjective priors for variances based on prior information is a difficult task. We therefore consider so-called automatic or default Bayes factors. These methods avoid the need for the user to specify priors by using information from the sample data. We discuss three automatic Bayes factors for testing variances. The first is a balanced Bayes factor with equal priors on all variances, where the priors are specified automatically using a small share of the information in the sample data. The second is the fractional Bayes factor, where a fraction of the likelihood is used for automatic prior specification. The third is an adjustment of the fractional Bayes factor such that the parsimony of inequality constrained hypotheses is properly taken into account. The results of a simulation study indicate that the balanced Bayes factor converges fastest to a true equality constrained hypothesis, whereas the adjusted fractional Bayes factor converges fastest to a true inequality constrained hypothesis.

4.1 Introduction

In comparing multiple independent populations, applied researchers commonly focus on the populations means, while treating the population variances as nuisance pa-

This chapter is under review at *Psychometrika* (revise and resubmit).

rameters. However, by disregarding the variances one runs the risk of overlooking crucial information in the data about the differences in the populations. In fact, there are often reasons to expect certain relations between the variances of independent populations. For example, Arden and Plomin (2006) expected boys to be more heterogeneous in their intelligence than girls. This expectation can be formalized in the inequality constrained hypothesis $H_1\colon \sigma_1^2 < \sigma_2^2$, where σ_1^2 and σ_2^2 denote the population variance of girls and boys, respectively. Potential competing hypotheses would be $H_0\colon \sigma_1^2 = \sigma_2^2$ and $H_2\colon \sigma_2^2 < \sigma_1^2$. In another study, Aunola, Leskinen, Lerkkanen, and Nurmi (2004) expected that the variance of students' mathematics abilities either increases or decreases across grades. These expectations can be translated into the two competing hypotheses $H_1\colon \sigma_1^2 < \cdots < \sigma_J^2$ and $H_2\colon \sigma_J^2 < \cdots < \sigma_1^2$, where σ_j^2 denotes the population variance in grade j and J is the number of grades to be compared. In an experiment one may expect variances in treatment groups to be larger than the variance in a control group because subjects may react differently to a certain treatment (e.g. Grissom, 2000). This suggests testing a hypothesis of the form $H_1\colon \sigma_1^2 < \sigma_2^2 = \sigma_3^2$, where σ_1^2 denotes the variance in the control group and σ_2^2 and σ_3^2 denote the variance in treatment groups 1 and 2, respectively. We could test H_1 against $H_2\colon \sigma_1^2 < \left(\sigma_2^2, \sigma_3^2\right)$ to determine whether there is evidence in favor of equal treatment group variances. The comma symbol in H_2 indicates that there is no constraint on the relation between σ_2^2 and σ_3^2. Another potential competing hypothesis would be the null hypothesis $H_0\colon \sigma_1^2 = \sigma_2^2 = \sigma_3^2$. In case there is just one treatment that is administered in two different intensities, one may expect that an intense treatment results in a larger variance than a mild treatment. This suggests testing the order constrained hypothesis $H_3\colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$, where group 2 undergoes mild treatment and group 3 intense treatment.

In this chapter we shall be interested in testing $T \geqslant 2$ hypotheses on the variances of $J \geqslant 2$ independent populations. The hypotheses are of the form

$$H_t\colon \boldsymbol{R}_t^E \boldsymbol{\sigma}^2 = \mathbf{0} \wedge \boldsymbol{R}_t^I \boldsymbol{\sigma}^2 > \mathbf{0}, \quad t = 1, \ldots, T, \tag{4.1}$$

where $\boldsymbol{\sigma}^2 = \begin{bmatrix}\sigma_1^2 & \cdots & \sigma_J^2\end{bmatrix}^T$ is a J-dimensional vector containing the population variances. Let q_t^E and q_t^I denote the number of equality and inequality constraints on the variances in $\boldsymbol{\sigma}^2$ under H_t, respectively. Then $\boldsymbol{R}_t^E$ ($\boldsymbol{R}_t^I$) is a $q_t^E \times J$ ($q_t^I \times J$) matrix containing the coefficients for the equality (inequality) constraints on the variances under H_t and $\mathbf{0} = \begin{bmatrix}0 & \cdots & 0\end{bmatrix}^T$ is a q_t^E-dimensional (q_t^I-dimensional) vector of zeroes. We shall consider tests where each row of $\boldsymbol{R}_t^E$ and $\boldsymbol{R}_t^I$ is a permutation of $\{-1, 1, 0, \ldots, 0\}$. Thus, we shall test constraints with equal coefficients for the variances (e.g. $\sigma_1^2 < \sigma_2^2$), but not constraints of the form $2\sigma_1^2 < \sigma_2^2$ or $\sigma_1^2 + \sigma_2^2 < \sigma_3^2$, for example. Note that the formulation in Equation (4.1) comprises many more hypotheses next to the classical null and alternative hypothesis, which are included as special cases. To our knowledge the multiple testing problem in Equation (4.1) has not yet been considered in the literature. This is quite surprising given the central role of variance components in the statistical sciences (see also Carroll, 2003).

In this chapter we shall consider the Bayes factor (Jeffreys, 1961; Kass & Raftery, 1995) for the testing problem formulated in Equation (4.1). The Bayes factor is a Bayesian hypothesis testing criterion that is becoming increasingly popular. It has a number of advantages over alternative approaches to hypothesis testing like null hypothesis significance testing by means of p-values and hypothesis testing by means

of information criteria like the AIC (Akaike, 1973) and the BIC (Schwarz, 1978): First, unlike p-values, Bayes factors are able to quantify the evidence in the data in favor of a hypothesis (including null hypotheses) relative to another hypothesis (Berger & Sellke, 1987; Wagenmakers, 2007). Second, using Bayes factors it is straightforward to simultaneously test multiple (non)nested hypotheses (Berger & Mortera, 1999). This property is not shared by p-values either. Third, Bayes factors are consistent in the sense that they converge to the true hypothesis as the sample size increases. This also holds for a true null hypothesis. The AIC, on the other hand, is not consistent (e.g. O'Hagan, 1995), and p-values are only consistent if the null hypothesis is false. Fourth, Bayes factors function as Occam's razor by automatically taking the parsimony of (in)equality constrained hypotheses as in Equation (4.1) into account. By contrast, p-values have no inherent mode of taking the parsimony of a hypothesis into account. The AIC and the BIC are able to incorporate the parsimony introduced by equality constraints, but not inequality constraints (Mulder et al., 2009). Consequently, they do not provide a solution to the testing problem in Equation (4.1).

Application of Bayes factors requires the specification of a prior distribution under every hypothesis to be tested. Often, however, prior information about the parameters is not available or a researcher would like to refrain from adding prior knowledge. For such situations researchers developed so-called automatic or default Bayes factors. These methods enable the computation of Bayes factors without having the user specify proper subjective priors. Automatic Bayes factors have been developed for various testing problems frequently encountered in practice: Rouder, Speckman, Sun, Morey, and Iverson (2009) proposed an automatic Bayesian t-test. Klugkist, Laudy, and Hoijtink (2005) presented an automatic Bayes factor for testing inequality constrained hypotheses on means in ANOVA models. Mulder, Hoijtink, and de Leeuw (2012) developed Bayes factors for testing (in)equality constraints on means and regression coefficients in multivariate normal linear regression models. Mulder (2016) applied the Bayes factor to the problem of testing order constrained hypotheses on correlations. Recently, Böing-Messing and Mulder (2016) developed automatic Bayes factors for testing constraints on the variances of two independent populations. In this chapter we extend their methods to the general case of testing variances of $J \geqslant 2$ independent populations. Note that the number of possible hypotheses increases considerably with the number of populations. For example, while in the $J = 2$ populations case there are four hypotheses ($H_0\colon \sigma_1^2 = \sigma_2^2$, $H_1\colon \sigma_1^2 < \sigma_2^2$, $H_2\colon \sigma_2^2 < \sigma_1^2$, and the unconstrained hypothesis $H_u\colon \sigma_1^2, \sigma_2^2$), using Equation (4.1) we can formulate dozens of hypotheses on the variances of $J = 3$ populations. The number of possible hypotheses quickly grows larger as the number of populations increases further. This underlines the importance of generalizing the existing methods to the $J \geqslant 2$ populations case.

In this chapter we will present three different automatic Bayes factors for testing (in)equality constrained hypotheses on variances as in Equation (4.1). The main idea of the three methods is to use a small share of the information in the sample data to automatically specify proper priors. Subsequently, the remaining share is used for hypothesis testing. This methodology avoids the need for the user to specify proper subjective priors based on prior information. Note that specifying subjective priors for variances under multiple (in)equality constrained hypotheses is a difficult task. In our experience, applied researchers find it difficult to formulate prior distributions that exactly capture their prior beliefs about the heterogeneity across populations.

Therefore automatic Bayesian methods are particularly useful in this context. The Bayes factors will be evaluated based on three criteria: the performance when testing nested inequality constrained hypotheses, information consistency, and large sample consistency.

The remainder of this chapter is structured as follows. In the next section we give a brief introduction to the Bayes factor. Following this, we develop the three automatic Bayes factors for testing variances. Next, we present the results of three numerical studies investigating the behavior of the Bayes factors when testing nested inequality constrained hypotheses, and whether the methods are information consistent and large sample consistent. After that we illustrate the practical utility of the Bayes factors by applying them to three exemplary data sets. We conclude the chapter with a discussion of the proposed methods.

4.2 The Bayes Factor

The Bayes factor is defined as the ratio of the marginal likelihoods under two competing hypotheses H_t and $H_{t'}$:

$$B_{tt'} = \frac{m_t(\boldsymbol{x})}{m_{t'}(\boldsymbol{x})}, \tag{4.2}$$

where $m_t(\boldsymbol{x})$ denotes the marginal likelihood under H_t as a function of the data $\boldsymbol{x} = \begin{bmatrix}\boldsymbol{x}_1^T & \cdots & \boldsymbol{x}_J^T\end{bmatrix}^T$. In this chapter we assume that the data $\boldsymbol{x}_j = \begin{bmatrix}x_{1j} & \cdots & x_{n_j j}\end{bmatrix}^T$ come from a normal population with mean μ_j and variance σ_j^2:

$$x_{ij} \overset{\text{i.i.d.}}{\sim} N\left(\mu_j, \sigma_j^2\right), \quad i = 1, \ldots, n_j, \; j = 1, \ldots, J. \tag{4.3}$$

Before we give the expression for the marginal likelihood under an (in)equality constrained hypothesis H_t as in Equation (4.1) we need to introduce some additional notation. Under a hypothesis H_t with q_t^E equality constraints and q_t^I inequality constraints on J population variances there are $K_t = J - q_t^E$ unique variances which we denote by $\boldsymbol{\sigma}_t^2 = \begin{bmatrix}\sigma_1^2 & \cdots & \sigma_{K_t}^2\end{bmatrix}^T$ (note that we omitted the hypothesis index t on the individual variances to simplify the notation). Furthermore, let J_k be the number of populations sharing the unique variance σ_k^2 and let μ_{k_j} denote the mean of the jth population sharing the unique variance σ_k^2, for $j = 1, \ldots, J_k$ and $k = 1, \ldots, K_t$. Similarly, let $\boldsymbol{x}_{k_j} = \begin{bmatrix}x_{1jk} & \cdots & x_{n_{k_j} jk}\end{bmatrix}^T$ be the vector of n_{k_j} observations from the jth population sharing the unique variance σ_k^2. If there are no equality constraints under H_t we shall omit the subscript j and write μ_k, $\boldsymbol{x}_k$, and n_k instead of μ_{k_1}, $\boldsymbol{x}_{k_1}$, and n_{k_1} to simplify the notation. In a similar manner, under the null hypothesis where there is just 1 unique variance we shall omit the subscript k and write σ^2, μ_j, $\boldsymbol{x}_j$, and n_j instead of σ_1^2, μ_{1_j}, $\boldsymbol{x}_{1_j}$, and n_{1_j}. Finally, we denote the admissible parameter space of the unique variances under H_t by Ω_t and the vector of the unconstrained population means by $\boldsymbol{\mu}$.

We illustrate the above notation by means of the hypothesis $H_1\colon \sigma_1^2 = \sigma_2^2 < \sigma_3^2$ on the variances of $J = 3$ populations. Under H_1 there is $q_1^E = 1$ equality constraint and $q_1^I = 1$ inequality constraint, resulting in $K_1 = 3 - 1 = 2$ unique variances denoted

by $\boldsymbol{\sigma}_1^2 = \begin{bmatrix}\sigma_1^2 & \sigma_2^2\end{bmatrix}^T$. Population 1 and 2, which have equal variances under H_1, share the unique variance σ_1^2 and population 3 has the unique variance σ_2^2. Consequently, the number of populations sharing the unique variances σ_1^2 and σ_2^2 is given by $J_1 = 2$ and $J_2 = 1$, respectively. Furthermore, $\boldsymbol{\mu} = \begin{bmatrix}\mu_{1_1} & \mu_{1_2} & \mu_{2_1}\end{bmatrix}^T$ is the vector of the means of population 1, 2, and 3. Similarly, $\boldsymbol{x} = \begin{bmatrix}\boldsymbol{x}_{1_1}^T & \boldsymbol{x}_{1_2}^T & \boldsymbol{x}_{2_1}^T\end{bmatrix}^T$ is the vector of the data from population 1, 2, and 3 with sample sizes of n_{1_1}, n_{1_2}, and n_{2_1}, respectively. Finally, the admissible parameter space of the unique variances under H_1 is given by $\Omega_1 = \left\{\boldsymbol{\sigma}_1^2 : \sigma_1^2 < \sigma_2^2\right\}$.

Using the notation introduced above, the marginal likelihood under an (in)equality constrained hypothesis H_t is given by

$$m_t(\boldsymbol{x}) = \int_{\Omega_t} \int_{\mathbb{R}^J} f_t\left(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) \pi_t\left(\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}_t^2. \tag{4.4}$$

The expression $f_t\left(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right)$ is the likelihood under H_t, which is given by

$$\begin{aligned} f_t\left(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) &= \prod_{k=1}^{K_t} \prod_{j=1}^{J_k} f\left(\boldsymbol{x}_{k_j}|\mu_{k_j}, \sigma_k^2\right) \mathbf{1}_{\Omega_t}\left(\boldsymbol{\sigma}_t^2\right) \\ &= \prod_{k=1}^{K_t} \prod_{j=1}^{J_k} \prod_{i=1}^{n_{k_j}} N\left(x_{ijk}|\mu_{k_j}, \sigma_k^2\right) \mathbf{1}_{\Omega_t}\left(\boldsymbol{\sigma}_t^2\right), \end{aligned} \tag{4.5}$$

where $\mathbf{1}_{\Omega_t}\left(\boldsymbol{\sigma}_t^2\right)$ is the indicator function which is 1 if $\boldsymbol{\sigma}_t^2 \in \Omega_t$ and 0 otherwise. The second component of the marginal likelihood in Equation (4.4) is $\pi_t\left(\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right)$, the prior distribution of the model parameters under H_t. Note that we may not use improper priors. This is because improper priors depend on unspecified constants. As a consequence, the Bayes factor would depend on a ratio of two unspecified constants (for details see, e.g., O'Hagan, 1995).

The marginal likelihood m_t quantifies how well an (in)equality constrained hypothesis H_t with prior π_t is able to predict the observed data (Jeffreys, 1961). Consequently, the Bayes factor $B_{tt'}$ quantifies how much better H_t is able to predict the data as compared to $H_{t'}$. The Bayes factor $B_{tt'}$ can thus be interpreted as a measure of relative evidence in the data in favor of H_t relative to $H_{t'}$.

To facilitate interpretation, one may compute the posterior probabilities of the hypotheses under investigation using the marginal likelihoods and the prior probabilities of the hypotheses $P(H_1), \ldots, P(H_T)$. The prior probabilities quantify the likelihood of the hypotheses before observing any data. A widely accepted default choice is to set equal prior probabilities $P(H_1) = \cdots = P(H_T) = 1/T$ (e.g. Berger & Mortera, 1999; Hoijtink, 2011; Mulder et al., 2012). After observing the data, the posterior probabilities of the hypotheses are obtained by updating the prior probabilities with the marginal likelihoods according to

$$P(H_t|\boldsymbol{x}) = \frac{m_t(\boldsymbol{x})P(H_t)}{\sum_{t'=1}^{T} m_{t'}(\boldsymbol{x})P(H_{t'})}, \tag{4.6}$$

for $t = 1, \ldots, T$. The resulting posterior probabilities $P(H_1|\boldsymbol{x}), \ldots, P(H_T|\boldsymbol{x})$ quantify the likelihood of the hypotheses after observing the data.

4.3 Automatic Bayes Factors

In the previous section we saw that in order to quantify the relative evidence in the data between the hypotheses of interest one needs to specify a proper prior for the free parameters under each hypothesis. However, specifying priors for the population variances under all hypotheses to be tested is a difficult task from a practical point of view. For this reason we shall focus on automatic Bayes factors which can be computed without needing to formulate proper subjective priors.

4.3.1 Balanced Bayes Factor

The main idea of the balanced Bayes factor (BBF) is to use information from the sample data to construct a proper prior in an automatic fashion such that it is balanced. We use the term "balanced" following Jeffreys (1961), who referred to an unconstrained prior for an effect as balanced if the prior probability of a positive effect is equal to the prior probability of a negative effect. The automatic prior in the BBF is based on a similar idea, namely, that every possible ordering of the population variances is equally likely a priori (similar as in Mulder, Hoijtink, and Klugkist (2010) for population means and regression coefficients). This balanced prior for the population variances contains minimal information and has a scale hyperparameter that is automatically determined by the sample data to avoid the need for subjective prior information. To obtain this balanced prior we proceed as follows. First, we fit a null model with a common variance to a small part of the sample data. The latter is obtained by taking a small fraction of the likelihood as suggested by O'Hagan (1995) in his fractional Bayes factor methodology. Next, we obtain the marginal posterior of the common variance based on this small fraction of the likelihood. We shall choose the fraction of the likelihood such that this marginal posterior contains minimal information (details will be discussed below). Finally, this posterior is used as prior for each unique variance under the constrained hypotheses. Note that under this prior different orderings of the variances are equally likely because every unique variance has the same prior.

The technical details of our approach to constructing the automatic balanced prior in the BBF are as follows. First, we assume $H_0\colon \sigma_1^2 = \ldots = \sigma_J^2 = \sigma^2$. We then obtain a proper posterior by updating the noninformative Jeffreys prior on $\boldsymbol{\mu}$ and σ^2 with a fraction of the likelihood under H_0:

$$\pi_0^B\left(\boldsymbol{\mu}, \sigma^2 | \boldsymbol{x}^{\boldsymbol{b}}\right) \propto \left(\prod_{j=1}^{J} f\left(\boldsymbol{x}_j | \mu_j, \sigma^2\right)^{b_j}\right) \pi_0^N\left(\boldsymbol{\mu}, \sigma^2\right), \tag{4.7}$$

where $\pi_0^N\left(\boldsymbol{\mu}, \sigma^2\right) \propto \sigma^{-2}$ is the Jeffreys prior under H_0, and we use the superscript B to refer to the BBF. The expression $f\left(\boldsymbol{x}_j | \mu_j, \sigma^2\right)^{b_j}$ denotes a fraction of the likelihood of the data from population j under H_0 (inspired by the fractional Bayes factor of O'Hagan, 1995). It is obtained by raising the likelihood of population j to the power of $b_j \in [0, 1]$. The exponent b_j is a population-specific fraction that controls how much information (in terms of the number of observations) is contained in the fraction of the likelihood of population j (Berger & Pericchi, 2001; De Santis & Spezzaferri, 2001). We use the notation $\boldsymbol{x}^{\boldsymbol{b}}$, where $\boldsymbol{b} = \begin{bmatrix} b_1 & \cdots & b_J \end{bmatrix}^T$, to indicate that the posterior in

Equation (4.7) contains a fraction of the information in the complete sample data. The larger the b's, the more information from the likelihood (i.e. from the sample data) is contained in the posterior.

In the next step we integrate $\boldsymbol{\mu}$ out of the joint posterior to obtain the marginal posterior of σ^2:

$$\pi_0^B\left(\sigma^2|\boldsymbol{x^b}\right) = \int_{\mathbb{R}^J} \pi_0^B\left(\boldsymbol{\mu}, \sigma^2|\boldsymbol{x^b}\right) d\boldsymbol{\mu} = \text{Inv-}\chi^2\left(\sigma^2|\nu, \tau^2\right), \tag{4.8}$$

where

$$\nu = \left(\sum_{j=1}^{J} b_j n_j\right) - J \quad \text{and} \quad \tau^2 = \frac{\sum_{j=1}^{J} b_j (n_j - 1) s_j^2}{\left(\sum_{j=1}^{J} b_j n_j\right) - J}. \tag{4.9}$$

Here Inv-$\chi^2\left(\nu, \tau^2\right)$ is the scaled inverse-χ^2 distribution with degrees of freedom parameter $\nu > 0$ and scale parameter $\tau^2 > 0$ (Gelman et al., 2004), and $s_j^2 = \frac{1}{n_j - 1}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_j)^2$ is the sample variance of $\boldsymbol{x}_j$.

We then define the prior on the unique variances $\boldsymbol{\sigma}_t^2 = \begin{bmatrix}\sigma_1^2 & \cdots & \sigma_{K_t}^2\end{bmatrix}^T$ under an (in)equality constrained hypothesis H_t as

$$\pi_t^B\left(\boldsymbol{\sigma}_t^2|\boldsymbol{x^b}\right) = \frac{1}{P^B\left(\boldsymbol{\sigma}_t^2 \in \Omega_t|\boldsymbol{x^b}\right)} \prod_{k=1}^{K_t} \pi_0^B\left(\sigma_k^2|\boldsymbol{x^b}\right) \mathbf{1}_{\Omega_t}\left(\boldsymbol{\sigma}_t^2\right), \tag{4.10}$$

where

$$P^B\left(\boldsymbol{\sigma}_t^2 \in \Omega_t|\boldsymbol{x^b}\right) = \int_{\Omega_t} \prod_{k=1}^{K_t} \pi_0^B\left(\sigma_k^2|\boldsymbol{x^b}\right) d\boldsymbol{\sigma}_t^2 \tag{4.11}$$

is the prior probability that the inequality constraints on the unique variances hold. In Equation (4.10) its inverse acts as a normalizing constant. The prior in Equation (4.10) is referred to as balanced because it implies that every possible ordering of variances is equally likely a priori. For example, under $H_1\colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ the prior probability P^B in Equation (4.11) equals 1/6 because all $3! = 6$ orderings of 3 variances are equally likely a priori.

The prior in Equation (4.10) must not be too vague or else Bartlett's phenomenon is induced (e.g. Bartlett, 1957; Jeffreys, 1961; Liang et al., 2008; Lindley, 1957). On the other hand, the prior should not be too informative either because then it would dominate the data. A widely accepted principle that provides a solution to this problem is to let the prior contain minimal information (e.g. Berger & Pericchi, 1996; O'Hagan, 1995; Spiegelhalter & Smith, 1982). We can make the scaled inverse-χ^2 prior in Equation (4.8) contain minimal information by setting the degrees of freedom to 1. This can be achieved by setting the fractions to $b_j = (1 + 1/J)/n_j$, for $j = 1, \ldots, J$. This gives us degrees of freedom of $\nu = \left(\sum_{j=1}^{J} b_j n_j\right) - J = \left(\sum_{j=1}^{J}(1 + 1/J)\right) - J = 1$ regardless of the sample sizes $n_1, \ldots, n_J$. Note that the scale parameter τ^2 in Equation (4.9) can be interpreted as a weighted average of sums of squares across all populations.

The unconstrained mean vector $\boldsymbol{\mu}$ is common under all hypotheses, which is why we use the noninformative Jeffreys prior $\pi^N(\boldsymbol{\mu}) = C$ for it (Jeffreys, 1961), where C

is an unspecified normalizing constant (see, e.g., O'Hagan, 1995). The joint balanced prior on the means and the variances under H_t is then given by

$$\pi_t^B\left(\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2 | \boldsymbol{x^b}\right) = \pi_t^B\left(\boldsymbol{\sigma}_t^2 | \boldsymbol{x^b}\right) \pi^N(\boldsymbol{\mu}). \tag{4.12}$$

Eventually, we define the marginal likelihood under a constrained hypothesis H_t based on the balanced prior as

$$m_t^B(\boldsymbol{x}, \boldsymbol{b}) = \int_{\Omega_t} \int_{\mathbb{R}^J} f_t\left(\boldsymbol{x} | \boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) \pi_t^B\left(\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2 | \boldsymbol{x^b}\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}_t^2. \tag{4.13}$$

After some algebra (see Appendix 4.A), this expression can be written as

$$\begin{aligned} &m_t^B(\boldsymbol{x}, \boldsymbol{b}) \\ &\quad = C \frac{P^B\left(\boldsymbol{\sigma}_t^2 \in \Omega_t | \boldsymbol{x}\right)}{P^B\left(\boldsymbol{\sigma}_t^2 \in \Omega_t | \boldsymbol{x^b}\right)} \left(\nu\tau^2\right)^{\frac{\nu K_t}{2}} \Gamma\left(\frac{\nu}{2}\right)^{-K_t} \pi^{-\frac{\sum_{k=1}^{K_t}\left(\left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k\right)}{2}} \left(\prod_{k=1}^{K_t} \prod_{j=1}^{J_k} n_{k_j}^{-\frac{1}{2}}\right) \\ &\qquad \prod_{k=1}^{K_t} \Gamma\left(\frac{\nu + \left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}{2}\right) \left(\nu\tau^2 + \sum_{j=1}^{J_k} \left(n_{k_j} - 1\right) s_{k_j}^2\right)^{-\frac{\nu + \left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}{2}}, \end{aligned} \tag{4.14}$$

where C is the unspecified normalizing constant from the Jeffreys prior on the means, $\Gamma(\cdot)$ is the gamma function, and $s_{k_j}^2$ is the sample variance of the data from the jth population sharing the unique variance σ_k^2. Furthermore,

$$\begin{aligned} &P^B\left(\boldsymbol{\sigma}_t^2 \in \Omega_t | \boldsymbol{x}\right) \\ &\quad = \int_{\Omega_t} \prod_{k=1}^{K_t} \text{Inv-}\chi^2\left(\sigma_k^2 \middle| \nu + \left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k, \frac{\nu\tau^2 + \sum_{j=1}^{J_k}\left(n_{k_j} - 1\right) s_{k_j}^2}{\nu + \left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}\right) d\boldsymbol{\sigma}_t^2 \end{aligned} \tag{4.15}$$

is the posterior probability that the inequality constraints on the unique variances hold. Note that the unspecified constant C cancels out in the computation of Bayes factors. The integral in Equation (4.15) cannot be computed analytically, but it can be approximated numerically using Monte Carlo methods (see Appendix 4.C).

At this point it is important to note that the BBF is scale invariant, that is, it does not depend on the scale of the outcome variable. This is an important property in theory as well as in practice. For example, in testing hypotheses on the variances of exam scores the evidence should not depend on whether exams are rated on a scale from 0 to 10 or 0 to 100. To show that the BBF is scale invariant we can proceed as follows. Let $w\boldsymbol{x}_{k_j}$ be the rescaled data of the jth group sharing the unique variance σ_k^2, where w is a constant. Then the sample variance of $w\boldsymbol{x}_{k_j}$ is given by $w^2 s_{k_j}^2$. If we substitute $s_{k_j}^2$ in Equation (4.14) with $w^2 s_{k_j}^2$, it can be shown that the marginal likelihood based on the rescaled data is equal to a hypothesis-independent constant times the marginal likelihood based on the original data. The hypothesis-independent constant cancels out in the computation of Bayes factors and posterior probabilities of the hypotheses. Thus, the Bayes factors and posterior probabilities based on the rescaled data are equal to those based on the original data, which shows that the BBF is scale invariant.

4.3.2 Fractional Bayes Factor

The next automatic Bayes factor we consider is the fractional Bayes factor (FBF) introduced by O'Hagan (1995). The FBF is a general method for Bayesian hypothesis testing and model selection. The main idea is to use a fraction of the information in the data to automatically specify a prior and to subsequently use the remaining fraction for hypothesis testing. A key difference from the BBF is that the automatic prior implied by the FBF is not balanced because it is concentrated around the likelihood. In this section we apply the FBF for the first time to the problem of testing hypotheses on $J \geqslant 2$ variances. Furthermore, we propose a generalization of the original FBF based on the use of population-specific fractions. This was also suggested by De Santis and Spezzaferri (2001) for testing means. We use the superscript F to refer to the FBF.

In the FBF approach the marginal likelihood under hypothesis H_t is defined as (O'Hagan, 1995)

$$m_t^F(\boldsymbol{x}, b) = \frac{\int_{\Omega_t} \int_{\mathbb{R}^J} f_t\left(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) \pi_t^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}_t^2}{\int_{\Omega_t} \int_{\mathbb{R}^J} f_t\left(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right)^b \pi_t^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}_t^2}. \tag{4.16}$$

Here $\pi_t^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right)$ is the noninformative Jeffreys prior on the population means and variances given by

$$\pi_t^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) = C_t \prod_{k=1}^{K_t} \sigma_k^{-2}\, \mathbf{1}_{\Omega_t}\left(\boldsymbol{\sigma}_t^2\right), \tag{4.17}$$

where C_t is an unspecified normalizing constant (see, e.g., O'Hagan, 1995). The expression $f_t\left(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right)^b$ is the likelihood under H_t to the power of b, a key part of the FBF methodology. The fraction b is a proportion that determines how much of the information in the likelihood (in terms of observations) is contained in $f_t\left(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right)^b$. Note that b is the same under all hypotheses. This is because b effectively divides the likelihood into a training fraction and a test fraction (see, e.g., Gilks, 1995), and the size of these two fractions should be constant across hypotheses.

Choosing the fraction b is a crucial step in the application of the FBF. A popular and widely accepted approach is setting $b = m_0/n$, where m_0 is the size of a minimal training sample and n is the sample size (e.g. Berger & Mortera, 1999; O'Hagan, 1995). This way the information in the data that is used for hypothesis testing is maximal. The use of different fractions for different parts of the likelihood has been recommended in the literature to avoid consistency issues (e.g. De Santis & Spezzaferri, 2001). We therefore use population-specific fractions $b_{k_j} = m_0/n_{k_j} = 2/n_{k_j}$, where $m_0 = 2$ because we need two observations from each population for the automatic prior under the unconstrained hypothesis to be proper. The fraction of the likelihood is then given by

$$f_t\left(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right)^{\boldsymbol{b}} = \prod_{k=1}^{K_t} \prod_{j=1}^{J_k} f\left(\boldsymbol{x}_{k_j}|\mu_{k_j}, \sigma_k^2\right)^{b_{k_j}} \mathbf{1}_{\Omega_t}\left(\boldsymbol{\sigma}_t^2\right), \tag{4.18}$$

where we slightly abuse notation by using the vector of population-specific fractions $\boldsymbol{b}$ as a superscript. Plugging the expression above into Equation (4.16), it can be shown

(see Appendix 4.B) that the marginal likelihood is equal to

$$m_t^F(\boldsymbol{x}, \boldsymbol{b}) = \frac{P^F\left(\boldsymbol{\sigma}_t^2 \in \Omega_t | \boldsymbol{x}\right)}{P^F\left(\boldsymbol{\sigma}_t^2 \in \Omega_t | \boldsymbol{x}^{\boldsymbol{b}}\right)} \pi^{-\frac{\sum_{k=1}^{K_t}\sum_{j=1}^{J_k}\left(1-b_{k_j}\right)n_{k_j}}{2}} \left(\prod_{k=1}^{K_t}\prod_{j=1}^{J_k} b_{k_j}^{\frac{1}{2}}\right) \prod_{k=1}^{K_t} \Gamma\left(\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}{2}\right) \Gamma\left(\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}{2}\right)^{-1} \left(\sum_{j=1}^{J_k}\left(n_{k_j}-1\right)s_{k_j}^2\right)^{-\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}} \left(\sum_{j=1}^{J_k} b_{k_j}\left(n_{k_j}-1\right)s_{k_j}^2\right)^{\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right)-J_k}{2}}, \tag{4.19}$$

where

$$P^F\left(\boldsymbol{\sigma}_t^2 \in \Omega_t | \boldsymbol{x}^{\boldsymbol{b}}\right) = \int_{\Omega_t} \prod_{k=1}^{K_t} \text{Inv-}\chi^2\left(\sigma_k^2 \,\middle|\, \left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k, \frac{\sum_{j=1}^{J_k} b_{k_j}\left(n_{k_j}-1\right)s_{k_j}^2}{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}\right) d\boldsymbol{\sigma}_t^2 \tag{4.20}$$

is the prior probability that the inequality constraints on the unique variances hold. The expression for the posterior probability that the inequality constraints hold is identical to Equation (4.20) with all b's equal to 1, that is,

$$P^F\left(\boldsymbol{\sigma}_t^2 \in \Omega_t | \boldsymbol{x}\right) = \int_{\Omega_t} \prod_{k=1}^{K_t} \text{Inv-}\chi^2\left(\sigma_k^2 \,\middle|\, \left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k, \frac{\sum_{j=1}^{J_k}\left(n_{k_j}-1\right)s_{k_j}^2}{\left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}\right) d\boldsymbol{\sigma}_t^2. \tag{4.21}$$

As for the BBF, the integrals in Equations (4.20) and (4.21) can be approximated using Monte Carlo methods (see Appendix 4.C). Furthermore, it can be shown that the FBF is scale invariant using the same approach as for the BBF (see Section 4.3.1).

4.3.3 Adjusted Fractional Bayes Factor

In testing equality constrained hypotheses, the FBF functions as Occam's razor by taking the parsimony introduced by equality constraints into account. However, it has been shown that the FBF may not function as Occam's razor when testing inequality constrained hypotheses (Böing-Messing & Mulder, 2016; Mulder, 2014b). If an inequality constrained hypothesis is strongly supported by the data, both the likelihood and the fraction of the likelihood are concentrated in the inequality constrained parameter space. As a result, the marginal likelihood under the inequality constrained hypothesis is approximately equal to the marginal likelihood under the unconstrained hypothesis. Consequently, the fractional Bayes factor for testing the inequality constrained hypothesis against the unconstrained hypothesis is approximately equal to 1 (as will be illustrated in the next section). Thus, the Bayes factor is indecisive even though the data strongly support the more parsimonious inequality constrained hypothesis. In this section we present an adjustment of the FBF such that it properly

takes the parsimony of inequality constrained hypotheses into account. We shall refer to this method as the adjusted fractional Bayes factor (aFBF). We use the superscript aF to refer to the aFBF.

In the aFBF approach we define the marginal likelihood under hypothesis H_t as

$$m_t^{aF}(\boldsymbol{x}, \boldsymbol{b}) = \frac{\int_{\Omega_t} \int_{\mathbb{R}^J} f_u\left(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) \pi_u^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}_t^2}{\int_{\Omega_t^a} \int_{\mathbb{R}^J} f_u\left(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right)^{\boldsymbol{b}} \pi_u^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}_t^2}. \tag{4.22}$$

where we use the same population-specific fractions $\boldsymbol{b}$ as in the FBF, that is, we set $b_{k_j} = 2/n_{k_j}$. The formulation above features two differences from the marginal likelihood in the FBF approach in Equation (4.16). First, in the denominator we integrate the variances over an adjusted parameter space Ω_t^a given by

$$\Omega_t^a = \left\{ \boldsymbol{\sigma}_t^2 \,\middle|\, \boldsymbol{R}_t^I \begin{bmatrix} a_1\sigma_1^2 & \cdots & a_{K_t}\sigma_{K_t}^2 \end{bmatrix}^T > \boldsymbol{0} \right\}, \tag{4.23}$$

where $a_1, \ldots, a_{K_t}$ are tuning parameters given by

$$a_k = \frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}{\sum_{j=1}^{J_k} b_{k_j}\left(n_{k_j} - 1\right) s_{k_j}^2}, \tag{4.24}$$

for $k = 1, \ldots, K_t$. The reason for this particular choice of tuning parameters will be explained later.

The second difference in Equation (4.22) is that we use the unconstrained likelihood and Jeffreys prior instead of the inequality constrained likelihood and Jeffreys prior under H_t. The unconstrained Jeffreys prior is given by $\pi_u^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) = C_{t,u} \prod_{k=1}^{K_t} \sigma_k^{-2}$. Using the unconstrained likelihood and Jeffreys prior is necessary to ensure that we integrate over the complete adjusted parameter space Ω_t^a in the denominator in Equation (4.22). If we use the unconstrained Jeffreys prior in the denominator we also need to use it in the numerator to make sure that the unspecified normalizing constant cancels out. Despite this adjustment, it is important to note that the numerator in the aFBF is still equal to the marginal likelihood under H_t based a noninformative improper prior, similar as in the original FBF.

The final expression for the marginal likelihood in the aFBF approach is identical to that of the FBF given in Equation (4.19), except that the prior probability that the inequality constraints hold is given by

$$\begin{aligned} &P^{aF}\left(\boldsymbol{\sigma}_t^2 \in \Omega_t^a \middle| \boldsymbol{x}^{\boldsymbol{b}}\right) \\ &\quad = \int_{\Omega_t^a} \prod_{k=1}^{K_t} \text{Inv-}\chi^2\left(\sigma_k^2 \,\middle|\, \left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k, \frac{\sum_{j=1}^{J_k} b_{k_j}\left(n_{k_j} - 1\right) s_{k_j}^2}{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}\right) d\boldsymbol{\sigma}_t^2 \\ &\quad = \int_{\Omega_t} \prod_{k=1}^{K_t} \text{Inv-}\chi^2\left(a_k\sigma_k^2 \,\middle|\, \left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k, a_k \frac{\sum_{j=1}^{J_k} b_{k_j}\left(n_{k_j} - 1\right) s_{k_j}^2}{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}\right) d\boldsymbol{\sigma}_t^2, \end{aligned} \tag{4.25}$$

where the two integrals are equal due to the mathematical result that if $\sigma^2 \sim \text{Inv-}\chi^2\left(\nu, \tau^2\right)$, then $a\sigma^2 \sim \text{Inv-}\chi^2\left(\nu, a\tau^2\right)$. As with the BBF and FBF, the

integrals above can be approximated using Monte Carlo methods (see Appendix 4.C). Furthermore, the aFBF is scale invariant for the same reason that the BBF and the FBF are.

Similar as in the FBF (cf. Equation (4.20)), the scale parameters of the scaled inverse-χ^2 distributions in the second line of Equation (4.25) are functions of the sample variances $s_{k_j}^2$. However, unlike in the FBF, the scale parameters also depend on the tuning parameters a_k. The idea behind the aFBF is to choose the tuning parameters such that scale parameters are independent of the sample variances. One such choice is the expression in Equation (4.24), which is just the inverse of the scale parameters in the first line of Equation (4.25). This particular choice for the tuning parameters gives us scale parameters equal to 1 for all variances. As a result, every possible ordering of variances is equally likely a priori (similar as in the BBF). For example, under $H_1\colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ we have $a_k\sigma_k^2|\boldsymbol{x}^{\boldsymbol{b}} \sim \text{Inv-}\chi^2(1,1)$, for $k = 1, 2, 3$, where the degrees of freedom are equal to $\nu_k = b_k n_k - 1 = 2/n_k \times n_k - 1 = 1$. Consequently, all 6 possible orderings of the 3 variances are equally likeliy a priori. In particular, the prior probability that the inequality constraints under H_1 hold is equal to $P^{aF}\left(a_1\sigma_1^2 < a_2\sigma_2^2 < a_3\sigma_3^2|\boldsymbol{x}^{\boldsymbol{b}}\right) = 1/6$.

4.4 Performance of the Bayes Factors

In this section we present the results of three numerical studies evaluating the automatic Bayes factors based on three criteria: the performance when testing nested inequality constrained hypotheses, information consistency, and large sample consistency.

4.4.1 Testing Nested Inequality Constrained Hypotheses

To illustrate the testing behavior of the three automatic Bayes factors for inequality constrained hypotheses in particular, we test the order constrained hypothesis $H_1\colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ against the unconstrained hypothesis $H_u\colon \sigma_1^2, \sigma_2^2, \sigma_3^2$. The top row of Figure 4.1 shows the BBF (solid line), the FBF (dashed line), and the aFBF (dotted line) of H_1 against H_u for common sample sizes of $n_1 = n_2 = n_3 = n = 5$ (left plot) and $n = 20$ (right plot) and sample variances of $\begin{bmatrix} s_1^2 & s_2^2 & s_3^2 \end{bmatrix}^T = \begin{bmatrix} 1 & s & s^2 \end{bmatrix}^T$. We let s^2 go from $\exp(0) = 1$ to $\exp(10) = 22026.47$. Thus, the larger s^2, the larger the size of the order effect. Note that setting $s_2^2 = s$ results in equal sample variance ratios of $s_2^2/s_1^2 = s_3^2/s_2^2 = s$. For the BBF we set $b_k = (1 + 1/3)/n$, whereas for the FBF and the aFBF we set $b_k = 2/n$.

Now, according to the Occam's razor principle H_1 should be favored over H_u if the constraints under H_1 are supported by the data since H_1 is more parsimonious than H_u (in the sense that the admissible parameter space under H_1 is a subset of the unconstrained space under H_u). It can be seen, however, that the FBF approaches 1 as s^2 grows very large. This means that the FBF is undecided about H_1 and H_u despite the fact that the data strongly support H_1, which suggests that the FBF does not function as Occam's razor in this case. This can be explained as follows. From the definition of the marginal likelihood in Equation (4.16) it follows that the FBF of

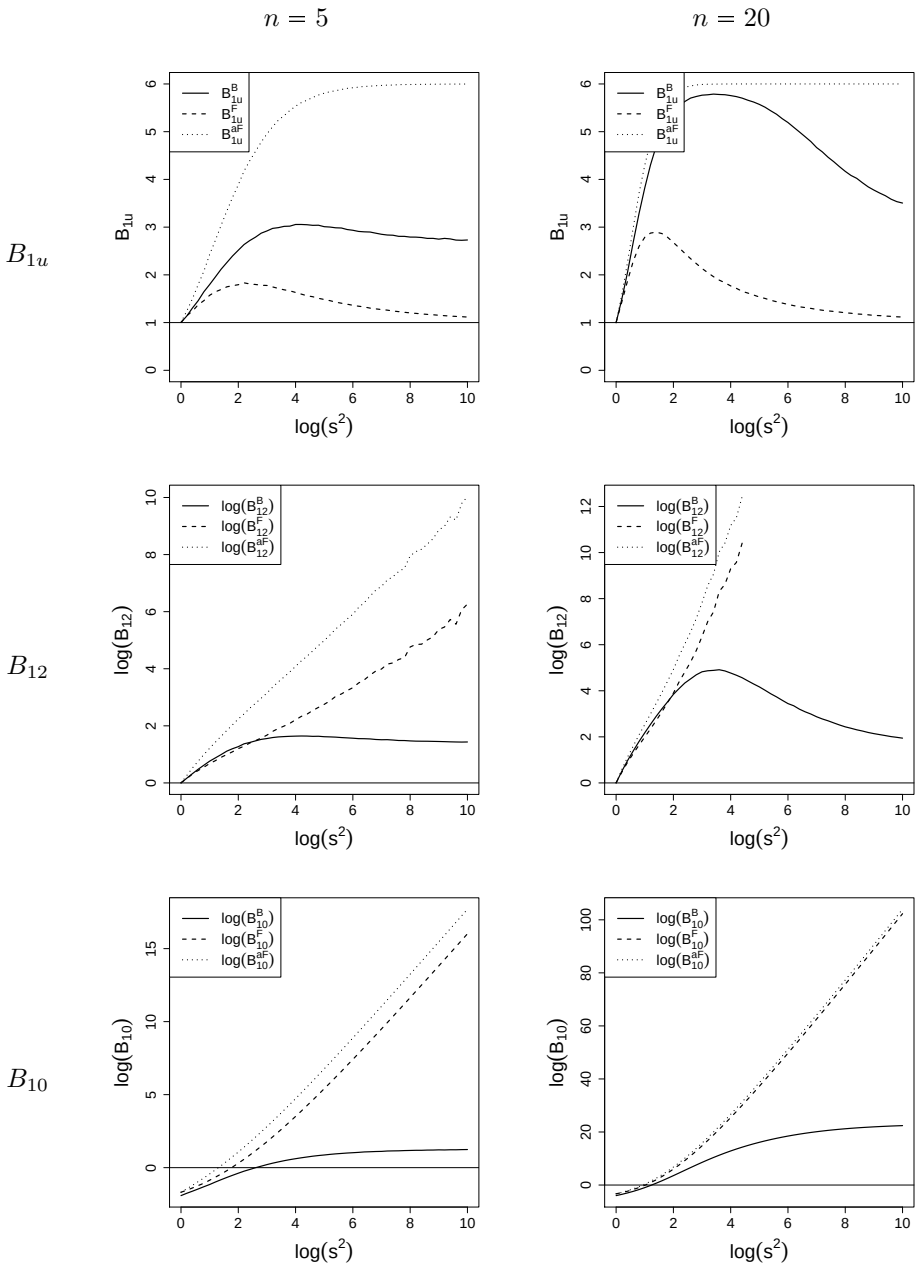


Figure 4.1: The BBF (solid line), FBF (dashed line), and aFBF (dotted line) testing $H_1\colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ against $H_u\colon \sigma_1^2, \sigma_2^2, \sigma_3^2$ (top row), $H_2\colon \neg H_1$ (middle row), and $H_0\colon \sigma_1^2 = \sigma_2^2 = \sigma_3^2$ (bottom row). The Bayes factors are plotted for common samples sizes of $n_1 = n_2 = n_3 = n = 5$ (left column) and $n = 20$ (right column) and sample variances of $\begin{bmatrix} s_1^2 & s_2^2 & s_3^2 \end{bmatrix}^T = \begin{bmatrix} 1 & s & s^2 \end{bmatrix}^T$, where $s^2 \in [\exp(0), \exp(10)]$. For the BBF we set $b_k = (1 + 1/3)/n$, whereas for the FBF and the aFBF we set $b_k = 2/n$.

H_1 against H_u can be written as

$$B^F_{1u} = \frac{P^F\left(\sigma_1^2 < \sigma_2^2 < \sigma_3^2 | \boldsymbol{x}\right)}{P^F\left(\sigma_1^2 < \sigma_2^2 < \sigma_3^2 | \boldsymbol{x^b}\right)} \rightarrow \frac{1}{1} = 1, \tag{4.26}$$

where $P^F\left(\sigma_1^2 < \sigma_2^2 < \sigma_3^2 | \boldsymbol{x}\right)$ and $P^F\left(\sigma_1^2 < \sigma_2^2 < \sigma_3^2 | \boldsymbol{x^b}\right)$ are the posterior and the prior probability that the inequality constraints under H_1 hold, respectively. Now, for very large effects both probabilities converge to 1, which results in a Bayes factor that converges to 1.

The BBF and the aFBF do not converge to 1 but to a value strictly larger than 1 as s^2 goes to infinity. The explanation for the BBF and the aFBF converging to constants greater than 1 is similar. First, similar to the FBF, it holds that

$$B^B_{1u} = \frac{P^B\left(\sigma_1^2 < \sigma_2^2 < \sigma_3^2 | \boldsymbol{x}\right)}{P^B\left(\sigma_1^2 < \sigma_2^2 < \sigma_3^2 | \boldsymbol{x^b}\right)} \rightarrow \frac{P^{B*}}{1/6} = 6 \times P^{B*} \tag{4.27}$$

and

$$B^{aF}_{1u} = \frac{P^{aF}\left(\sigma_1^2 < \sigma_2^2 < \sigma_3^2 | \boldsymbol{x}\right)}{P^{aF}\left(a_1\sigma_1^2 < a_2\sigma_2^2 < a_3\sigma_3^2 | \boldsymbol{x^b}\right)} \rightarrow \frac{1}{1/6} = 6. \tag{4.28}$$

For the BBF, the posterior probability converges to $P^{B*} \approx 0.45$ for $n = 5$ and $P^{B*} \approx 0.50$ for $n = 20$ as the effect size increases (it does not converge to 1 due to prior shrinkage). The prior probability always equals 1/6 since in the BBF approach the prior is the product of three identical distributions, such that each of the 6 possible orderings of the 3 variances is equally likely a priori. Consequently, the BBF converges to $6 \times P^{B*} \approx 2.69$ for $n = 5$ and $6 \times P^{B*} \approx 3.00$ for $n = 20$ as the effect size increases. In the aFBF approach the posterior probability goes to 1 as the effect size increases, and the tuning parameters a_1, a_2, a_3 adapt to the sample sizes and sample variances such that the prior probability always equals 1/6. As a result, the aFBF converges to 6. Thus, contrary to the FBF, the BBF and the aFBF function as Occam's razor by favoring the more parsimonious inequality constrained hypothesis H_1 over the unconstrained hypothesis H_u if the former is strongly supported by the data.

4.4.2 Information Consistency

A Bayes factor for an unconstrained hypothesis against the null hypothesis is called information consistent if it goes to infinity as the effect size goes to infinity, while keeping the sample size fixed. The Bayes factor is called information inconsistent if it converges to a constant in the limit. A well-known example of information inconsistency is the Bayes factor based on Zellner's g-prior (Zellner, 1986; Berger & Pericchi, 2001). Information (in)consistency when testing inequality constrained hypotheses was first considered by Mulder (2014a) for testing means. To our knowledge information (in)consistency has never been investigated when testing variances. Here we evaluate information consistency for two different tests on variances: (i) testing an inequality constrained hypothesis against its complement and (ii) testing an inequality constrained hypothesis against the null hypothesis. We will call the Bayes factor B_{12} (B_{10}) of an inequality constrained hypothesis $H_1\colon \boldsymbol{R}_1^I\boldsymbol{\sigma}^2 > \boldsymbol{0}$ against $H_2\colon \neg H_1$ ($H_0\colon \sigma_1^2 = \cdots = \sigma_J^2$) information consistent if $B_{12} \rightarrow \infty$ ($B_{10} \rightarrow \infty$) as each element

in $\hat{\boldsymbol{\xi}} = \boldsymbol{R}_1^I \begin{bmatrix} \log\left(\hat{\sigma}_1^2\right) & \cdots & \log\left(\hat{\sigma}_J^2\right) \end{bmatrix}^T$ goes to infinity, while keeping the sample size fixed. If the Bayes factor converges to a constant $B_{12}^* < \infty$ ($B_{10}^* < \infty$) instead, then it is referred to as information inconsistent.

First, we investigate information consistency when testing an inequality constrained hypothesis against its complement. The middle row of Figure 4.1 shows the logarithm of the BBF, the FBF, and the aFBF of $H_1 \colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ against $H_2 \colon \neg H_1$ as $\hat{\boldsymbol{\xi}} = \begin{bmatrix} \log\left(s_2^2/s_1^2\right) & \log\left(s_3^2/s_2^2\right) \end{bmatrix}^T$ increases from $\begin{bmatrix} 0 & 0 \end{bmatrix}^T$ to $\begin{bmatrix} 5 & 5 \end{bmatrix}^T$. The results indicate that the FBF and the aFBF are information consistent, whereas the BBF is information inconsistent. As the effect size increases, the BBF converges to a constant $B_{12}^{B^*} \approx \exp(1.40) = 4.05$ for $n = 5$ and $B_{12}^{B^*} \approx \exp(1.61) = 5.00$ for $n = 20$. This behavior of the BBF can be explained by the fact that the posterior probability that the inequality constraints under H_1 hold converges to $P^{B^*} \approx 0.45$ for $n = 5$ and $P^{B^*} \approx 0.50$ for $n = 20$ (as was found in Section 4.4.1), which implies that the probability that the inequality constraints under H_1 do not hold (as is stated in H_2) converges to $1 - P^{B^*} \approx 0.55$ for $n = 5$ and $1 - P^{B^*} \approx 0.50$ for $n = 20$. The BBF of H_1 against H_2 thus converges to $B_{12}^{B^*} = B_{1u}^{B^*} / B_{2u}^{B^*} \approx \frac{P^{B^*}}{1/6} \Big/ \frac{1 - P^{B^*}}{5/6}$, which equals 4.05 for $n = 5$ and 4.99 for $n = 20$. Note that H_1 is more parsimonious than H_2 because H_1 covers 1/6 of the unconstrained parameter space while H_2 covers 5/6 of the unconstrained space. The results show that the aFBF indicates stronger evidence in favor of the more parsimonious hypothesis H_1 than the FBF. This again illustrates that the aFBF functions as Occam's razor whereas the FBF does not.

Next, we investigate information consistency when testing the order constrained hypothesis $H_1 \colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ against the null hypothesis $H_0 \colon \sigma_1^2 = \sigma_2^2 = \sigma_3^2$. The bottom row of Figure 4.1 shows the logarithm of the Bayes factor B_{10} as a function of the effect size s^2. The results indicate that the FBF and the aFBF are information consistent since for these two Bayes factors the evidence in favor of H_1 goes to infinity as the size of the order effect increases. The BBF, on the other hand, does not show information consistent behavior in this case either. Again, the aFBF indicates stronger evidence in favor of H_1 than the FBF.

The inconsistent behavior of the BBF illustrates that it may not provide a good quantification of the relative evidence in the data between (in)equality constrained hypotheses in the case of small samples and large effects.

4.4.3 Large Sample Consistency

We next present the results of a simulation study investigating the performance of the three automatic Bayes factors as a function of the sample size. In particular, we are interested in whether the Bayes factors show consistent behavior in the sense that the posterior probability of the true hypothesis converges to 1 as the sample size increases.

Simulation Setup

In this simulation we tested hypotheses on the variances of $J = 4$ populations. We used a simulation design with two factors. The first factor is the pattern of the population variances. We considered five different variance patterns, referred to as null pattern, order pattern, mixed pattern, near order pattern, and reverse order

Table 4.1: Overview of population variances used in the simulation study.

Pattern	σ_1^2	σ_2^2	σ_3^2	σ_4^2
Null	1.00	1.00	1.00	1.00
Order	1.00	1.36	1.84	2.50
Mixed	1.00	1.00	1.58	2.50
Near order	1.36	1.00	1.84	2.50
Reverse order	2.50	1.84	1.36	1.00

pattern. An overview of the variance patterns is given in Table 4.1. In the null pattern we set all population variances equal to 1. In the order pattern the variances follow an increasing order. The magnitude of the order effect is given by the ratio of the largest to the smallest population variance. Empirical findings indicate that a ratio of $\sigma_4^2/\sigma_1^2 = 2.50$ can be considered a medium effect in the psychological sciences (see, e.g., Ruscio & Roche, 2012, Table 2). We determined σ_2^2 and σ_3^2 such that the ratio of adjacent population variances is constant, that is, $\sigma_2^2/\sigma_1^2 = \sigma_3^2/\sigma_2^2 = \sigma_4^2/\sigma_3^2 = 1.36$. In the mixed pattern the structure of the variances is $\sigma_1^2 = \sigma_2^2 < \sigma_3^2 < \sigma_4^2$. The near order pattern is similar to the order pattern with the difference that the variances of populations 1 and 2 are interchanged. Finally, in the reverse order pattern the variances are ordered as $\sigma_4^2 < \sigma_3^2 < \sigma_2^2 < \sigma_1^2$. The second factor in our simulation study is the size of the samples drawn from the four populations. We drew samples of common size $n_1 = \cdots = n_4 = n \in \{5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000\}$.

In each of the $5 \times 11 = 55$ conditions we drew 1000 samples $\boldsymbol{x}^{(m)} = \begin{bmatrix} \boldsymbol{x}_1^{(m)} & \cdots & \boldsymbol{x}_J^{(m)} \end{bmatrix}^T$, $m = 1, \ldots, 1000$. Here $\boldsymbol{x}_j^{(m)} = \begin{bmatrix} x_{1j}^{(m)} & \cdots & x_{n_j j}^{(m)} \end{bmatrix}^T$, where $x_{ij}^{(m)}$ is distributed as in Equation (4.3). We specified the population variances according to Table 4.1 and set all population means equal to 0 (note that the three Bayes factors do not depend on the sample means, cf. Equations (4.14), (4.19), and (4.25)). In each of the 1000 samples per condition we tested four hypotheses using the three different Bayes factors:

$$\begin{aligned} H_0 &: \sigma_1^2 = \cdots = \sigma_4^2, \\ H_1 &: \sigma_1^2 < \cdots < \sigma_4^2, \\ H_2 &: \sigma_1^2 = \sigma_2^2 < \sigma_3^2 < \sigma_4^2, \\ H_3 &: \neg\,(H_0 \vee H_1 \vee H_2). \end{aligned} \tag{4.29}$$

Here H_3 is the complement which comprises all possible hypotheses except H_0, H_1, and H_2. Note that the marginal likelihood under H_3 is equal to the marginal likelihood under the hypothesis $H_4 \colon \neg H_1$ because the probability of the event that two or more variances are exactly equal is 0. Furthermore, note that for the near order and the reverse order pattern the true hypothesis is contained in the complement H_3 (cf. Table 4.1). In each sample we then used the marginal likelihoods under all four hypotheses to compute the posterior probability of the true hypothesis H_t as $P\left(H_t|\boldsymbol{x}^{(m)}\right) = m_t\left(\boldsymbol{x}^{(m)}\right) \big/ \sum_{t'=0}^{3} m_{t'}\left(\boldsymbol{x}^{(m)}\right)$, where we assumed equal prior probabilities of the hypotheses. Eventually, we computed the expected posterior probability

of H_t as $\bar{P}(H_t|\boldsymbol{x}) = \frac{1}{1000}\sum_{m=1}^{1000} P\left(H_t|\boldsymbol{x}^{(m)}\right)$.

Results

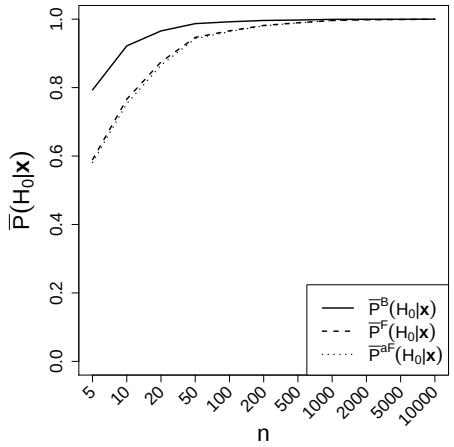
Figure 4.2 shows the simulation results for the five variance patterns. The plots show the expected posterior probability of the true hypothesis H_t, $t = 0, 1, 2, 3$, as a function of the common sample size n for the BBF (solid line), the FBF (dashed line), and the aFBF (dotted line). Overall, the results indicate that the three Bayes factors are consistent: As the sample size increases, the expected posterior probability of the true hypothesis approaches 1. It appears that the BBF converges fastest to a true hypothesis if two or more population variances are equal (see the null and the mixed pattern in Figures 4.2a and 4.2c, respectively), whereas the FBF and the aFBF converge fastest to the true hypotheses if none of the population variances are equal (see the order patterns in Figures 4.2b, 4.2d, and 4.2e). Furthermore, it can be seen that the FBF and the aFBF behave similarly. The FBF converges slightly faster to a true null hypothesis (see Figure 4.2a), whereas the aFBF converges somewhat faster to a true inequality constrained hypothesis (see Figures 4.2b and 4.2c).

Under the null pattern (Figure 4.2a), sample sizes of 10 (BBF) and 20 (FBF, aFBF) result in posterior probabilities of the true null hypothesis H_0 of at least 0.8. Under the order pattern (Figure 4.2b) we need considerably larger samples sizes of 1000 (BBF) and 500 (FBF, aFBF) to obtain posterior probabilities of the true order constrained hypothesis H_1 of at least 0.8. Under the mixed pattern (Figure 4.2c), sample sizes of 100 (BBF) and 200 (FBF, aFBF) are sufficient for reaching values of at least 0.8 for the true mixed hypothesis H_2. Under the near order pattern (Figure 4.2d) we need sample sizes of 1000 (BBF, FBF, aFBF), whereas under the reverse order pattern (Figure 4.2e) considerably smaller sample sizes of 200 (BBF) and 100 (FBF, aFBF) result in posterior probabilities of the true complement hypothesis H_3 of at least 0.8.

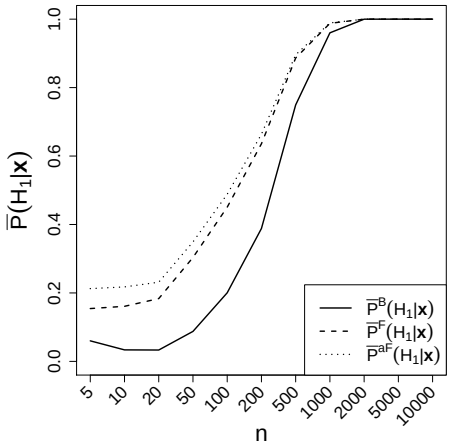
4.5 Example Applications

4.5.1 Example 1: Data From Weerahandi (1995)

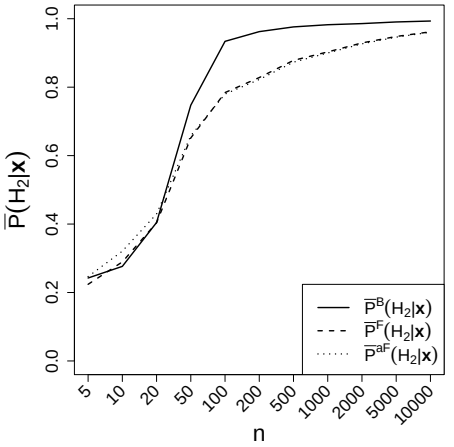
Weerahandi (1995) considered a hypothetical study with four treatment groups. The sample sizes and sample variances are shown in Table 4.2 (Example 1). In practice such a decreasing pattern of sample variances could emerge, for example, if the groups receive a new drug in a decreasing dosage. Patients may respond quite differently to a new drug, especially if the dosage is high. As a result, the variance is larger in groups receiving higher dosages. To check this we shall test the following constrained hypotheses on the group variances: $H_0\colon \sigma_1^2 = \cdots = \sigma_4^2$, $H_1\colon \sigma_4^2 < \cdots < \sigma_1^2$, and $H_2\colon \neg\,(H_0 \vee H_1)$. The results for the three Bayes factors are shown in Table 4.3 (Example 1). The posterior probabilities of the hypotheses were computed assuming equal prior probabilities. It can be seen that the BBF favors H_0. The FBF and the aFBF, on the other hand, favor H_1, with the aFBF indicating considerably weaker evidence in favor of H_0 and H_2. Overall, the results are in line with the findings of the simulation study, where the BBF provides stronger evidence in favor of the null hypothesis, whereas the FBF and the aFBF yield stronger evidence in favor of



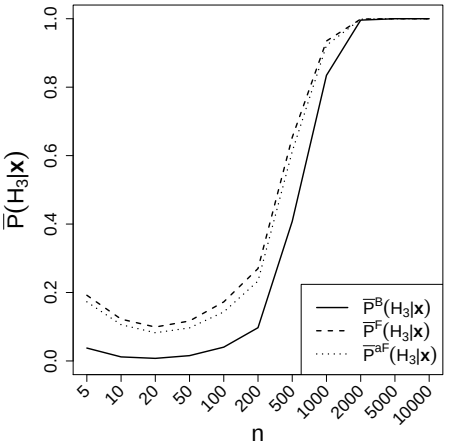
(a) Null



(b) Order

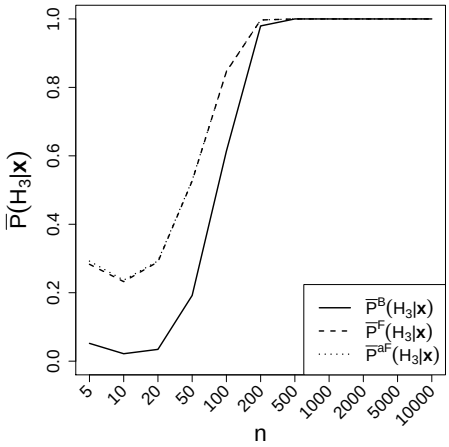


(c) Mixed



(d) Near order

Figure 4.2: Results of a simulation study comparing the performance of the three automatic Bayes factors in testing variances of four populations. We considered five different patterns of the population variances: (a) $\sigma_1^2 = \cdots = \sigma_4^2$, (b) $\sigma_1^2 < \cdots < \sigma_4^2$, (c) $\sigma_1^2 = \sigma_2^2 < \sigma_3^2 < \sigma_4^2$, (d) $\sigma_2^2 < \sigma_1^2 < \sigma_3^2 < \sigma_4^2$, and (e) $\sigma_4^2 < \cdots < \sigma_1^2$. In each of the five cases we drew 1000 samples of size $n_1 = \cdots = n_4 = n$. In each sample we then tested four hypotheses: $H_0\colon \sigma_1^2 = \cdots = \sigma_4^2$, $H_1\colon \sigma_1^2 < \cdots < \sigma_4^2$, $H_2\colon \sigma_1^2 = \sigma_2^2 < \sigma_3^2 < \sigma_4^2$, and $H_3\colon \neg(H_0 \vee H_1 \vee H_2)$. Eventually, we computed the expected posterior probability of the true hypothesis $\bar{P}(H_t|\boldsymbol{x})$ across the 1000 samples. The plots show $\bar{P}(H_t|\boldsymbol{x})$ as a function of the common sample size n.



(e) Reverse order

Figure 4.2 (continued)

inequality constrained hypotheses. The fact that the FBF and the aFBF support the order constrained hypothesis H_1 despite the small sample sizes is due to the large effect size of $s_1^2/s_4^2 = 3.61/0.30 = 11.93$. Comparing the logarithm of this effect size with the results in the bottom row of Figure 4.1 indicates that the preference of the BBF for the null hypothesis may be a result of information inconsistency: From the plots it can be seen that for an effect size of $\log(11.93) = 2.48$ the BBF already shows information inconsistent behavior. This suggests relying on the results of the FBF or the aFBF, which indicate evidence in favor of the order constrained hypothesis H_1 stating that the heterogeneity of the responses increases with the dosage.

Table 4.2: Samples sizes and sample variances for three examples.

Example	Group		n	s^2
Example 1	1	Treatment 1	6	3.61
	2	Treatment 2	8	2.89
	3	Treatment 3	5	0.79
	4	Treatment 4	7	0.30
Example 2	1	Controls	17	15.52
	2	Tourette's patients	17	20.07
	3	ADHD patients	17	38.81
Example 3	1	male leader, appointed at random	30	3.46
	2	female leader, appointed at random	30	1.32
	3	male leader, appointed on ability	30	3.20
	4	female leader, appointed on ability	30	2.10

Table 4.3: Results for three examples. The posterior probabilities of the hypotheses were computed assuming equal prior probabilities. In some cases the posterior probabilities do not sum to 1 due to rounding.

Example	Bayes factor	$P(H_0\|\boldsymbol{x})$	$P(H_1\|\boldsymbol{x})$	$P(H_2\|\boldsymbol{x})$	$P(H_3\|\boldsymbol{x})$
Example 1	BBF	0.74	0.23	0.04	—
	FBF	0.12	0.72	0.17	—
	aFBF	0.04	0.91	0.05	—
Example 2	BBF	0.35	0.48	0.14	0.03
	FBF	0.28	0.40	0.25	0.07
	aFBF	0.24	0.43	0.28	0.06
Example 3	BBF	0.37	0.62	0.00	—
	FBF	0.16	0.82	0.03	—
	aFBF	0.12	0.86	0.02	—

4.5.2 Example 2: Attentional Performances of Tourette's and ADHD Patients

Silverstein, Como, Palumbo, West, and Osborn (1995) conducted a study comparing attentional performances of 17 Tourette's and 17 ADHD patients with those of a group of 17 controls. Participants were shown 120 strings of 12 letters. Each string contained either a T or an F at a random position, the remaining 11 letters were random letters other than T and F. Each string was presented for 55 milliseconds. After each presentation, participants had to indicate as quickly as possible whether the string contained a T or an F. After completion of the 120 strings the accuracy of the respondents was computed as the percentage of correct answers. Table 4.2 (Example 2) shows the sample variances of the accuracies in the three groups.

Psychological research has frequently found ADHD patients to be more heterogeneous in their attentional performances than unaffected controls (see, e.g., Kofler et al., 2013; Russell et al., 2006). The heterogeneity of attentional performances of Tourette's patients as compared to unaffected controls is less well-documented. Given this information, we shall test the following hypotheses to investigate whether there is evidence that Tourette's patients (group 2) are as heterogeneous in their attentional performances as either unaffected controls (group 1) or ADHD patients (group 3): $H_1\colon \sigma_1^2 = \sigma_2^2 < \sigma_3^2$ and $H_2\colon \sigma_1^2 < \sigma_2^2 = \sigma_3^2$. We shall compare H_1 and H_2 to the competing hypotheses $H_0\colon \sigma_1^2 = \sigma_2^2 = \sigma_3^2$ and $H_3\colon \neg\,(H_0 \vee H_1 \vee H_2)$. The results are shown in Table 4.3 (Example 2). It can be seen that the three automatic Bayes factors produce similar results. In particular, the three Bayes factors favor H_1, which states that Tourette's patients are as heterogeneous as unaffected controls, and both groups are less heterogeneous than ADHD patients. However, while we can rule out the complement H_3, the posterior probabilities indicate some evidence in favor of H_0 and H_2. It can be seen that the aFBF provides somewhat stronger evidence in favor of the inequality constrained hypotheses than the FBF. This behavior was also observed in the numerical studies in Section 4.4.

4.5.3 Example 3: Influence of Group Leaders

Lucas (2003) investigated the influence of group leaders on subordinate group members. He was interested in whether a leader's influence depends on the leader's gender and the way the leader was appointed. The author considered two types of appointment: Either the leader was chosen at random or based on ability. Lucas conducted a 2×2 factorial experiment with 30 participants in each condition. Influence of the group leader was measured as the number of times (in 10 trials) that a participant changed his/her opinion to match the group leader's opinion. Table 4.2 (Example 3) shows the sample variances of the counts in the four experimental groups.

Research on gender differences suggests that the variability is greater for male leaders than for female leaders (e.g. Lehre et al., 2009). Due to a lack of theoretical underpinning we shall assume that there is no effect of appointment type. These expectations correspond to the hypothesis $H_1\colon \sigma_2^2 = \sigma_4^2 < \sigma_1^2 = \sigma_3^2$. We shall test H_1 against the competing hypotheses $H_0\colon \sigma_1^2 = \cdots = \sigma_4^2$ and $H_2\colon \neg\,(H_0 \vee H_1)$. The results of the test are shown in Table 4.3 (Example 3). It can be seen that H_1 receives strongest support from all three automatic Bayes factors. While there is some evidence in favor of H_0 (especially for the BBF), the complement H_2 can be ruled out given posterior probabilities close to 0.

4.6 Conclusion

In this chapter we presented three automatic Bayes factors for testing (in)equality constrained hypotheses on variances. We first introduced the balanced Bayes factor, which is based on identical automatic priors for the unique variances under each hypothesis. The hyperparameters of this prior are determined automatically using information from the sample data. The second Bayes factor is the fractional Bayes factor of O'Hagan (1995), which we derived for testing (in)equality constrained hypotheses on variances. We proposed a generalization of the fractional approach using population-specific fractions instead of a common fraction. The third Bayes factor we presented is an adjustment of the fractional Bayes factor such that the parsimony of inequality constrained hypotheses is incorporated. The three Bayes factors are fully automatic in the sense that there is no need for the user to specify priors under all hypotheses to be tested. Instead, the user only needs to provide the sample sizes and sample variances.

Results of numerical studies indicated that the FBF does not properly function as Occam's razor because it does not favor a parsimonious inequality constrained hypothesis that is supported by the data over the more complex unconstrained hypothesis. The BBF and the aFBF, on the other hand, always behaved as Occam's razor when testing inequality constrained hypotheses. Furthermore, the results indicated that the BBF is information inconsistent when testing (in)equality constrained hypotheses on variances. The FBF and the aFBF, on the other hand, showed information consistent behavior. Based on these findings we recommend the aFBF for quantifying the relative evidence in the data between multiple constrained hypotheses on variances when prior information about the magnitude of the effects is unavailable.

4.A Computation of $m_t^B(\boldsymbol{x}, \boldsymbol{b})$

The final expression for the marginal likelihood under an (in)equality constrained hypothesis H_t in the balanced Bayes factor can be derived as follows:

$$
\begin{aligned}
&m_t^B(\boldsymbol{x}, \boldsymbol{b}) \\
&= \int_{\Omega_t} \int_{\mathbb{R}^J} f_t\left(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) \pi_t^B\left(\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2|\boldsymbol{x}^{\boldsymbol{b}}\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}_t^2 \\
&= \int_{\Omega_t} \int_{\mathbb{R}^J} \left(\prod_{k=1}^{K_t} \prod_{j=1}^{J_k} f\left(\boldsymbol{x}_{k_j}|\mu_{k_j}, \sigma_k^2\right) \right) \\
&\qquad\qquad C \frac{1}{P^B\left(\boldsymbol{\sigma}_t^2 \in \Omega_t|\boldsymbol{x}^{\boldsymbol{b}}\right)} \prod_{k=1}^{K_t} \text{Inv-}\chi^2\left(\sigma_k^2|\nu, \tau^2\right) \mathbf{1}_{\Omega_t}\left(\boldsymbol{\sigma}_t^2\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}_t^2 \\
&= C \frac{1}{P^B\left(\boldsymbol{\sigma}_t^2 \in \Omega_t|\boldsymbol{x}^{\boldsymbol{b}}\right)} \int_{\Omega_t} \prod_{k=1}^{K_t} \left(\frac{\nu\tau^2}{2}\right)^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)^{-1} \left(\sigma_k^2\right)^{-\left(\frac{\nu}{2}+1\right)} \exp\left(-\frac{\nu\tau^2}{2\sigma_k^2}\right) \\
&\quad \prod_{j=1}^{J_k} \int_{\mathbb{R}} \left(\sigma_k^2 2\pi\right)^{-\frac{n_{k_j}}{2}} \exp\left(-\frac{1}{2\sigma_k^2}\left(\left(n_{k_j}-1\right)s_{k_j}^2 + n_{k_j}\left(\bar{x}_{k_j}-\mu_{k_j}\right)^2\right)\right) d\mu_{k_j}\, d\boldsymbol{\sigma}_t^2 \\
&= C \frac{1}{P^B\left(\boldsymbol{\sigma}_t^2 \in \Omega_t|\boldsymbol{x}^{\boldsymbol{b}}\right)} \left(\frac{\nu\tau^2}{2}\right)^{\frac{\nu K_t}{2}} \Gamma\left(\frac{\nu}{2}\right)^{-K_t} \int_{\Omega_t} \prod_{k=1}^{K_t} \left(\sigma_k^2\right)^{-\left(\frac{\nu}{2}+1\right)} \exp\left(-\frac{\nu\tau^2}{2\sigma_k^2}\right) \\
&\qquad\qquad \prod_{j=1}^{J_k} n_{k_j}^{-\frac{1}{2}} \left(\sigma_k^2 2\pi\right)^{-\frac{n_{k_j}-1}{2}} \exp\left(-\frac{\left(n_{k_j}-1\right)s_{k_j}^2}{2\sigma_k^2}\right) d\boldsymbol{\sigma}_t^2 \\
&= C \frac{1}{P^B\left(\boldsymbol{\sigma}_t^2 \in \Omega_t|\boldsymbol{x}^{\boldsymbol{b}}\right)} \left(\frac{\nu\tau^2}{2}\right)^{\frac{\nu K_t}{2}} \Gamma\left(\frac{\nu}{2}\right)^{-K_t} (2\pi)^{-\frac{\sum_{k=1}^{K_t}\left(\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k\right)}{2}} \left(\prod_{k=1}^{K_t}\prod_{j=1}^{J_k} n_{k_j}^{-\frac{1}{2}}\right) \\
&\qquad \int_{\Omega_t} \prod_{k=1}^{K_t} \left(\sigma_k^2\right)^{-\left(\frac{\nu+\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}+1\right)} \exp\left(-\frac{\nu\tau^2 + \sum_{j=1}^{J_k}\left(n_{k_j}-1\right)s_{k_j}^2}{2\sigma_k^2}\right) d\boldsymbol{\sigma}_t^2 \\
&= C \frac{1}{P^B\left(\boldsymbol{\sigma}_t^2 \in \Omega_t|\boldsymbol{x}^{\boldsymbol{b}}\right)} \left(\nu\tau^2\right)^{\frac{\nu K_t}{2}} \Gamma\left(\frac{\nu}{2}\right)^{-K_t} \pi^{-\frac{\sum_{k=1}^{K_t}\left(\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k\right)}{2}} \left(\prod_{k=1}^{K_t}\prod_{j=1}^{J_k} n_{k_j}^{-\frac{1}{2}}\right) \\
&\qquad \left(\prod_{k=1}^{K_t} \Gamma\left(\frac{\nu + \left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}{2}\right) \left(\nu\tau^2 + \sum_{j=1}^{J_k}\left(n_{k_j}-1\right)s_{k_j}^2\right)^{-\frac{\nu+\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}} \right) \\
&\qquad\qquad \int_{\Omega_t} \prod_{k=1}^{K_t} \text{Inv-}\chi^2\left(\sigma_k^2 \middle| \nu + \left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k, \frac{\nu\tau^2 + \sum_{j=1}^{J_k}\left(n_{k_j}-1\right)s_{k_j}^2}{\nu + \left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}\right) d\boldsymbol{\sigma}_t^2
\end{aligned}
$$

$$= C \frac{P^B\left(\boldsymbol{\sigma}_t^2 \in \Omega_t | \boldsymbol{x}\right)}{P^B\left(\boldsymbol{\sigma}_t^2 \in \Omega_t | \boldsymbol{x}^{\boldsymbol{b}}\right)} \left(\nu\tau^2\right)^{\frac{\nu K_t}{2}} \Gamma\left(\frac{\nu}{2}\right)^{-K_t} \pi^{-\frac{\sum_{k=1}^{K_t}\left(\left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k\right)}{2}} \left(\prod_{k=1}^{K_t}\prod_{j=1}^{J_k} n_{k_j}^{-\frac{1}{2}}\right)$$
$$\prod_{k=1}^{K_t} \Gamma\left(\frac{\nu + \left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}{2}\right) \left(\nu\tau^2 + \sum_{j=1}^{J_k}\left(n_{k_j} - 1\right) s_{k_j}^2\right)^{-\frac{\nu + \left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}{2}}, \tag{4.30}$$

where in the third line we may drop the indicator function because the integration region for the variances is already restricted to Ω_t, and the integrand in the fifth line is a product of kernels of scaled inverse-χ^2 distributions with degrees of freedom parameters $\nu_k = \nu + \left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k$ and scale parameters $\tau_k^2 = \frac{\nu\tau^2 + \sum_{j=1}^{J_k}\left(n_{k_j}-1\right)s_{k_j}^2}{\nu + \left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}$.

4.B Computation of $m_t^F(\boldsymbol{x}, \boldsymbol{b})$

In the fractional Bayes factor the marginal likelihood under an (in)equality constrained hypothesis H_t is defined as

$$m_t^F(\boldsymbol{x}, \boldsymbol{b}) = \frac{\int_{\Omega_t}\int_{\mathbb{R}^J} f_t\left(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) \pi_t^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}_t^2}{\int_{\Omega_t}\int_{\mathbb{R}^J} f_t\left(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right)^{\boldsymbol{b}} \pi_t^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}_t^2} = \frac{m_t^N(\boldsymbol{x})}{m_t^N(\boldsymbol{x}^{\boldsymbol{b}})}. \tag{4.31}$$

We first derive the denominator:

$$\begin{aligned}
&m_t^N(\boldsymbol{x}^{\boldsymbol{b}}) \\
&\quad= \int_{\Omega_t}\int_{\mathbb{R}^J} \left(\prod_{k=1}^{K_t}\prod_{j=1}^{J_k} f\left(\boldsymbol{x}_{k_j}|\mu_{k_j}, \sigma_k^2\right)^{b_{k_j}}\right) C_t \prod_{k=1}^{K_t} \sigma_k^{-2}\, \mathbf{1}_{\Omega_t}\left(\boldsymbol{\sigma}_t^2\right)\, d\boldsymbol{\mu}\, d\boldsymbol{\sigma}_t^2 \\
&\quad= C_t \int_{\Omega_t} \prod_{k=1}^{K_t} \sigma_k^{-2} \prod_{j=1}^{J_k} \int_{\mathbb{R}} \left(\sigma_k^2 2\pi\right)^{-\frac{b_{k_j} n_{k_j}}{2}} \\
&\qquad\qquad \exp\left(-\frac{b_{k_j}}{2\sigma_k^2}\left(\left(n_{k_j} - 1\right) s_{k_j}^2 + n_{k_j}\left(\bar{x}_{k_j} - \mu_{k_j}\right)^2\right)\right) d\mu_{k_j}\, d\boldsymbol{\sigma}_t^2 \\
&\quad= C_t \int_{\Omega_t} \prod_{k=1}^{K_t} \sigma_k^{-2} \prod_{j=1}^{J_k} \left(b_{k_j} n_{k_j}\right)^{-\frac{1}{2}} \left(\sigma_k^2 2\pi\right)^{-\frac{b_{k_j} n_{k_j} - 1}{2}} \exp\left(-\frac{b_{k_j}\left(n_{k_j} - 1\right) s_{k_j}^2}{2\sigma_k^2}\right) d\boldsymbol{\sigma}_t^2
\end{aligned}$$

$$
\begin{aligned}
&= C_t\,(2\pi)^{-\frac{\sum_{k=1}^{K_t}\left(\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right)-J_k\right)}{2}} \left(\prod_{k=1}^{K_t}\prod_{j=1}^{J_k}\left(b_{k_j} n_{k_j}\right)^{-\frac{1}{2}}\right) \\
&\qquad \int_{\Omega_t} \prod_{k=1}^{K_t}\left(\sigma_k^2\right)^{-\left(\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right)-J_k}{2}+1\right)} \exp\left(-\frac{\sum_{j=1}^{J_k} b_{k_j}\left(n_{k_j}-1\right)s_{k_j}^2}{2\sigma_k^2}\right) d\boldsymbol{\sigma}_t^2 \\
&= C_t\,\pi^{-\frac{\sum_{k=1}^{K_t}\left(\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right)-J_k\right)}{2}} \left(\prod_{k=1}^{K_t}\prod_{j=1}^{J_k}\left(b_{k_j} n_{k_j}\right)^{-\frac{1}{2}}\right) \\
&\qquad \left(\prod_{k=1}^{K_t}\Gamma\left(\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right)-J_k}{2}\right)\left(\sum_{j=1}^{J_k} b_{k_j}\left(n_{k_j}-1\right)s_{k_j}^2\right)^{-\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right)-J_k}{2}}\right) \\
&\qquad\qquad \int_{\Omega_t}\prod_{k=1}^{K_t}\text{Inv-}\chi^2\left(\sigma_k^2\,\middle|\,\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right)-J_k, \frac{\sum_{j=1}^{J_k} b_{k_j}\left(n_{k_j}-1\right)s_{k_j}^2}{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right)-J_k}\right) d\boldsymbol{\sigma}_t^2 \\
&= C_t\,\pi^{-\frac{\sum_{k=1}^{K_t}\left(\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right)-J_k\right)}{2}} \left(\prod_{k=1}^{K_t}\prod_{j=1}^{J_k}\left(b_{k_j} n_{k_j}\right)^{-\frac{1}{2}}\right) \\
&\qquad \left(\prod_{k=1}^{K_t}\Gamma\left(\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right)-J_k}{2}\right)\left(\sum_{j=1}^{J_k} b_{k_j}\left(n_{k_j}-1\right)s_{k_j}^2\right)^{-\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right)-J_k}{2}}\right) \\
&\qquad\qquad P^F\left(\boldsymbol{\sigma}_t^2 \in \Omega_t \middle| \boldsymbol{x}^{\boldsymbol{b}}\right).
\end{aligned}
\tag{4.32}
$$

The expression for the numerator in Equation (4.31) is identical to the final expression in Equation (4.32) with all b's equal to 1, that is,

$$
\begin{aligned}
m_t^N(\boldsymbol{x}) &= C_t\,\pi^{-\frac{\sum_{k=1}^{K_t}\left(\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k\right)}{2}} \left(\prod_{k=1}^{K_t}\prod_{j=1}^{J_k} n_{k_j}^{-\frac{1}{2}}\right) \\
&\qquad \left(\prod_{k=1}^{K_t}\Gamma\left(\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}\right)\left(\sum_{j=1}^{J_k}\left(n_{k_j}-1\right)s_{k_j}^2\right)^{-\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}}\right) \\
&\qquad\qquad P^F\left(\boldsymbol{\sigma}_t^2 \in \Omega_t \middle| \boldsymbol{x}\right),
\end{aligned}
\tag{4.33}
$$

where

$$
P^F\left(\boldsymbol{\sigma}_t^2 \in \Omega_t \middle| \boldsymbol{x}\right) = \int_{\Omega_t}\prod_{k=1}^{K_t}\text{Inv-}\chi^2\left(\sigma_k^2\,\middle|\,\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k, \frac{\sum_{j=1}^{J_k}\left(n_{k_j}-1\right)s_{k_j}^2}{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}\right) d\boldsymbol{\sigma}_t^2. \tag{4.34}
$$

The final expression for the marginal likelihood in Equation (4.31) is then given by

$$\begin{aligned} m_t^F(\boldsymbol{x}, \boldsymbol{b}) &= \frac{m_t^N(\boldsymbol{x})}{m_t^N(\boldsymbol{x}^{\boldsymbol{b}})} \\ &= \frac{P^F\left(\boldsymbol{\sigma}_t^2 \in \Omega_t | \boldsymbol{x}\right)}{P^F\left(\boldsymbol{\sigma}_t^2 \in \Omega_t | \boldsymbol{x}^{\boldsymbol{b}}\right)} \pi^{-\frac{\sum_{k=1}^{K_t}\sum_{j=1}^{J_k}\left(1-b_{k_j}\right)n_{k_j}}{2}} \left(\prod_{k=1}^{K_t}\prod_{j=1}^{J_k} b_{k_j}^{\frac{1}{2}}\right) \\ &\quad \prod_{k=1}^{K_t} \Gamma\left(\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}{2}\right) \Gamma\left(\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}{2}\right)^{-1} \\ &\quad \left(\sum_{j=1}^{J_k}\left(n_{k_j}-1\right)s_{k_j}^2\right)^{-\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}} \left(\sum_{j=1}^{J_k} b_{k_j}\left(n_{k_j}-1\right)s_{k_j}^2\right)^{\frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right)-J_k}{2}}. \end{aligned} \tag{4.35}$$

4.C Computing the Probability That $\boldsymbol{\sigma}_t^2 \in \Omega_t$

The integrals in Equations (4.15), (4.20), and (4.25) can be approximated numerically using the following Monte Carlo approach. For the BBF and the FBF, we first sample $\sigma_k^{2(s)} \sim \text{Inv-}\chi^2\left(\nu_k, \tau_k^2\right)$, for $k = 1, \ldots, K_t$, where $\sigma_k^{2(s)}$ is the sth draw from $\text{Inv-}\chi^2\left(\nu_k, \tau_k^2\right)$, for $s = 1, \ldots, S$, and ν_k and τ_k^2 are as in Equations (4.15) and (4.20), respectively. An approximation of the probability that the inequality constraints on the unique variances hold is then given by the proportion of draws that fall in Ω_t, that is,

$$P\left(\boldsymbol{\sigma}_t^2 \in \Omega_t\right) \approx \frac{1}{S}\sum_{s=1}^{S} \mathbf{1}_{\Omega_t}\left(\boldsymbol{\sigma}_t^{2(s)}\right), \tag{4.36}$$

where $\boldsymbol{\sigma}_t^{2(s)} = \begin{bmatrix}\sigma_1^{2(s)} & \cdots & \sigma_{K_t}^{2(s)}\end{bmatrix}^T$, and $\mathbf{1}_{\Omega_t}\left(\boldsymbol{\sigma}_t^{2(s)}\right)$ is the indicator function which is 1 if $\boldsymbol{\sigma}_t^{2(s)} \in \Omega_t$ and 0 otherwise.

For the aFBF, let $\phi_k = a_k\sigma_k^2$. We then proceed analogously to the BBF and the FBF: First, we sample $\phi_k^{(s)} \sim \text{Inv-}\chi^2\left(\nu_k, \tau_k^2\right)$, for $k = 1, \ldots, K_t$ and $s = 1, \ldots, S$, where ν_k and τ_k^2 are as in the second row of Equation (4.25). Then

$$P^{aF}\left(\boldsymbol{\sigma}_t^2 \in \Omega_t^a | \boldsymbol{x}^{\boldsymbol{b}}\right) = P^{aF}\left(\boldsymbol{\phi}_t \in \Omega_t | \boldsymbol{x}^{\boldsymbol{b}}\right) \approx \frac{1}{S}\sum_{s=1}^{S} \mathbf{1}_{\Omega_t}\left(\boldsymbol{\phi}_t^{(s)}\right), \tag{4.37}$$

where $\boldsymbol{\phi}_t = \begin{bmatrix}\phi_1 & \cdots & \phi_{K_t}\end{bmatrix}^T$ and $\boldsymbol{\phi}_t^{(s)} = \begin{bmatrix}\phi_1^{(s)} & \cdots & \phi_{K_t}^{(s)}\end{bmatrix}^T$.

Chapter 5

Bayes Factors for Testing Inequality Constrained Hypotheses on Variances of Dependent Observations

Abstract

In statistical practice variances are commonly treated as nuisance parameters. Often, however, there are patterns of interest not only in the (conditional) means, but also in the variances. In fact, there are often reasons to expect a certain pattern of the variances of dependent observations. For example, in a repeated measures study one may expect the variability of the outcome to increase over time because subjects react differently to a certain treatment. Such expectations can be formulated as inequality constrained hypotheses on the variances of the dependent observations. Currently, no methods exist for testing such hypotheses in a direct manner. In this chapter we develop a Bayes factor for testing inequality constrained hypotheses on the variances in the multivariate normal linear model, which is often used to model dependent observations. Application of Bayes factors requires specification of a prior distribution under every hypothesis to be tested. We make use of the encompassing prior approach in which priors under inequality constrained hypotheses are formulated as truncations of the prior under the unconstrained hypothesis. Our method is fully automatic in the sense that it does not require specification of subjective priors.

5.1 Introduction

In analyzing their data researchers commonly focus on measures of central tendency such as means and regression coefficients. Measures such as variances that capture the heterogeneity of observations are usually regarded as nuisance parameters. However, in order to fully understand the patterns that are present in the data it is of vital importance to carefully model and interpret the variability of the observations (e.g.

Carroll, 2003). In fact, the heterogeneity of the observations can be a core aspect of a study (e.g. Aunola et al., 2004; Hultsch et al., 2002; Lehre et al., 2009). In this chapter we focus on the heterogeneity of dependent observations.

There are often reasons to expect certain patterns in the heterogeneity. For example, in a repeated measures study investigating the effect of a certain treatment one may expect the variability of the outcome to increase over time because subjects may respond differently to the treatment (e.g. Böing-Messing, van Assen, Hofman, Hoijtink, & Mulder, 2017). Such an expectation can be formulated as an inequality constrained hypothesis of the form $H_1\colon \sigma_1^2 < \cdots < \sigma_p^2$, where p is the number of measurement occasions and σ_j^2 is the variance of the jth measurement, for $j = 1, \ldots, p$. To test H_1 we need a hypothesis it can be compared to. A natural competitor is the complement of H_1 given by $H_2\colon \neg\, \sigma_1^2 < \cdots < \sigma_p^2$, which may also be written as $H_2\colon \neg H_1$ in short. The complement entails all possible hypotheses on the p variances except H_1. In another example, Aunola et al. (2004) hypothesized that the variability of math performances either increases or decreases across grades. An increase might occur because children with high mathematical potential improve their performances faster than children with low mathematical potential. A possible reason for a decrease is that systematic instruction at school helps children with low mathematical potential catch up. These two competing expectations can be expressed as inequality constrained hypotheses $H_1\colon \sigma_1^2 < \cdots < \sigma_p^2$ and $H_2\colon \sigma_p^2 < \cdots < \sigma_1^2$, where p is the number of grades and σ_j^2 is the variance in grade j, for $j = 1, \ldots, p$. Another conceivable competitor is the complement of H_1 and H_2 given by $H_3\colon \neg \left(\sigma_1^2 < \cdots < \sigma_p^2 \vee \sigma_p^2 < \cdots < \sigma_1^2\right)$.

In this chapter we use the multivariate normal linear model with $p \times p$ covariance matrix $\boldsymbol{\Sigma}$ to model the dependent observations. We shall be concerned with testing $T \geqslant 2$ inequality constrained hypotheses on the variances $\boldsymbol{\sigma}^2 = \begin{bmatrix}\sigma_1^2 & \cdots & \sigma_p^2\end{bmatrix}^T$ of the dependent observations located on the main diagonal of $\boldsymbol{\Sigma}$. In general, the hypotheses are of the form

$$H_t\colon \boldsymbol{R}_t\boldsymbol{\sigma}^2 > \mathbf{0}, \quad t = 1, \ldots, T, \tag{5.1}$$

where $\boldsymbol{R}_t$ is a $q_t \times p$ matrix containing the coefficients for the q_t inequality constraints on the variances under H_t and $\mathbf{0} = \begin{bmatrix}0 & \cdots & 0\end{bmatrix}^T$ is a q_t-dimensional vector of zeroes. We shall test hypotheses for which each row of $\boldsymbol{R}_t$ is a permutation of $\{-1, 1, 0, \ldots, 0\}$. That is, we consider hypotheses with equal coefficients on the variances (e.g. $H_1\colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2$, for which $\boldsymbol{R}_1 = \left[\begin{smallmatrix}-1 & 1 & 0\\ 0 & -1 & 1\end{smallmatrix}\right]$). We do not impose any constraints on the covariances and regression coefficients in the multivariate normal linear model, which is why these parameters are omitted from the formulation of the hypotheses in Equation (5.1). Note that we reverse common statistical practice by focusing on the variances, while treating the regression coefficients as nuisance parameters.

In this chapter we shall take a Bayesian approach to the testing problem in Equation (5.1) using the Bayes factor (Jeffreys, 1961; Kass & Raftery, 1995), which is a Bayesian hypothesis testing and model selection criterion. Bayes factors have been developed for a variety of testing problems frequently encountered in practice. For example, Klugkist, Laudy, and Hoijtink (2005) developed a Bayes factor for testing inequality constrained hypotheses on the means in AN(C)OVA models. Mulder (2014) proposed Bayes factors for testing inequality constrained hypotheses on means and regression coefficients in the multivariate normal linear model. Gu, Mulder, Deković,

and Hoijtink (2014) developed an approximate Bayesian procedure for testing inequality constrained hypotheses on regression coefficients in a structural equation modeling framework. Mulder (2016) proposed Bayes factors for testing inequality constrained hypotheses on correlations. Recently, Böing-Messing et al. (2017) applied the Bayes factor to the problem of testing inequality constrained hypotheses on the variances of independent observations. In this chapter we extend this literature by developing Bayes factors for testing inequality constrained hypotheses on variances of dependent observations.

The Bayes factor is particularly suited for our problem of testing $T \geqslant 2$ inequality constrained hypotheses on the variances of dependent observations for a number of reasons. First, using Bayes factors it is straightforward to simultaneously test multiple (non)nested hypotheses. Second, Bayes factors are able to quantify the evidence in favor of the hypotheses under investigation. Third, our Bayes factor is consistent in the sense that it is guaranteed to select the true hypothesis as the sample size goes to infinity (this will be explained later). Fourth, Bayes factors have an inherent Occam's razor mechanism that automatically takes the parsimony introduced by inequality constraints into account when evaluating the hypotheses (e.g. Kass & Raftery, 1995). On the other hand, classical information criteria such as the AIC (Akaike, 1973) and the BIC (Schwarz, 1978) are not suitable for testing hypotheses formulated according to Equation (5.1) for two reasons: First, contrary to the BIC, the AIC is not consistent (e.g. O'Hagan, 1995). Second, neither the AIC nor the BIC are able to take the parsimony of inequality constrained hypotheses into account. This is because both the AIC and the BIC measure the complexity of a hypothesis by the number of unknown parameters. This, however, is not a suitable measure when testing inequality constrained hypotheses. For example, under both the inequality constrained hypothesis $H_1 \colon \sigma_1^2 < \cdots < \sigma_p^2$ and the unconstrained hypothesis $H_u \colon \sigma_1^2, \ldots, \sigma_p^2 > 0$ there are p unknown variance parameters. However, H_1 is more parsimonious than H_u because the admissible parameter space under H_1 is smaller in the sense that it is a subset of the unconstrained parameter space under H_u.

To apply Bayes factors, a prior distribution needs to be specified under every hypothesis to be tested. The prior distributions contain the information about the model parameters that is available before observing the data. In this chapter we use the encompassing prior approach (Berger & Mortera, 1999; Klugkist et al., 2005) to specify priors under inequality constrained hypotheses when prior information about the magnitude of the variances is absent. We proceed as follows. First, we propose an unconstrained proper prior for the covariance matrix in the multivariate normal linear model based on the following considerations: If prior information is absent, then all possible orderings of the variances should be equally likely a priori and the contribution of the prior to the posterior should be negligible (i.e. the prior should be vague). We then formulate priors under competing inequality constrained hypotheses as truncations of the unconstrained prior in the respective inequality constrained parameter spaces. For the regression coefficients we shall use a noninformative improper prior. We may so do so since the regression coefficients are nuisance parameters that are common under all hypotheses, which means that the undefined constants in the improper prior cancel out when computing the Bayes factors (e.g. Jeffreys, 1961).

The remainder of this chapter is structured as follows. In the next section we describe the multivariate normal linear model that is used throughout this chapter

and we propose an unconstrained prior on the model parameters. Following this, we give a brief introduction to the Bayes factor. We then develop a Bayes factor for testing inequality constrained hypotheses on the variances in the multivariate normal linear model using the encompassing prior approach. Next, we present results of a simulation study investigating how the Bayes factor depends on the sample size when different hypotheses are true. Subsequently, we illustrate the use of the Bayes factor by applying it to an empirical data set about reading recognition in children. We conclude the chapter with a discussion of our approach to testing inequality constrained hypotheses on the variances of dependent observations.

5.2 Model and Unconstrained Prior

In this chapter we shall be concerned with testing inequality constrained hypotheses on the variances in the multivariate normal linear model

$$\boldsymbol{y}_i = \boldsymbol{B}^T\boldsymbol{x}_i + \boldsymbol{e}_i, \quad i = 1, \ldots, n, \tag{5.2}$$

where $\boldsymbol{y}_i = \begin{bmatrix} y_{i1} & \cdots & y_{ip} \end{bmatrix}^T$ is a p-dimensional vector of correlated dependent variables, $\boldsymbol{x}_i = \begin{bmatrix} 1 & x_{i1} & \cdots & x_{i,k-1} \end{bmatrix}^T$ is a k-dimensional vector of independent variables, and $\boldsymbol{e}_i \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. The unknown parameters in this model are

$$\boldsymbol{B} = \begin{bmatrix} \beta_{01} & \cdots & \beta_{0p} \\ \vdots & \ddots & \vdots \\ \beta_{k-1,1} & \cdots & \beta_{k-1,p} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_p^2 \end{bmatrix}. \tag{5.3}$$

We shall be interested in testing inequality constrained hypotheses on the variances on the main diagonal of $\boldsymbol{\Sigma}$, where the hypotheses are formulated according to Equation (5.1). Note that there are no restrictions on the regression coefficients in $\boldsymbol{B}$ and the covariances in $\boldsymbol{\Sigma}$.

The model in Equation (5.2) can be written more compactly as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}, \tag{5.4}$$

where $\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}_1 & \cdots & \boldsymbol{y}_n \end{bmatrix}^T$, $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \end{bmatrix}^T$, and $\boldsymbol{E} = \begin{bmatrix} \boldsymbol{e}_1 & \cdots & \boldsymbol{e}_n \end{bmatrix}^T$. The unconstrained likelihood for n observations is then given by

$$\begin{aligned} f_u(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{B}, \boldsymbol{\Sigma}) &= (2\pi)^{-np/2} \left|\boldsymbol{\Sigma}\right|^{-n/2} \exp\left(-\tfrac{1}{2} \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})\right)\right) \\ &= (2\pi)^{-np/2} \left|\boldsymbol{\Sigma}\right|^{-n/2} \exp\left(-\tfrac{1}{2} \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S})\right) \\ &\qquad \exp\left(-\tfrac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \left(\boldsymbol{\Sigma} \otimes (\boldsymbol{X}^T\boldsymbol{X})^{-1}\right)^{-1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right), \end{aligned} \tag{5.5}$$

where $\otimes$ denotes the Kronecker product, $\boldsymbol{\beta}$ is the vectorization of $\boldsymbol{B}$, $\hat{\boldsymbol{\beta}}$ is the vectorization of the least squares estimate $\hat{\boldsymbol{B}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$, and

$$\boldsymbol{S} = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{B}})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{B}}), \tag{5.6}$$

is the error sum of squares and cross products matrix.

A common choice for the unconstrained prior on the covariance matrix $\boldsymbol{\Sigma}$ is a conjugate inverse-Wishart prior:

$$\pi_u(\boldsymbol{\Sigma}) = W_p^{-1}(\boldsymbol{\Sigma}|\boldsymbol{\Lambda}, \nu), \tag{5.7}$$

where $\boldsymbol{\Lambda} > 0$ is the $p \times p$ prior scale matrix ($\boldsymbol{\Lambda} > 0$ means that $\boldsymbol{\Lambda}$ must be positive definite) and $\nu > p - 1$ are the prior degrees of freedom. We do not test hypotheses on any elements of the coefficient matrix $\boldsymbol{B}$. This means that $\boldsymbol{B}$ is a nuisance parameter that is common under all inequality constrained hypotheses to be tested on the variances. In this case we may use the standard noninformative Jeffreys prior $\pi^N(\boldsymbol{B}) = C$ for the coefficient matrix, where C is an unspecified normalizing constant that cancels out in the computation of Bayes factors (e.g. Jeffreys, 1961; O'Hagan, 1995). The joint unconstrained prior on $\boldsymbol{\Sigma}$ and $\boldsymbol{B}$ is then given by

$$\pi_u(\boldsymbol{B}, \boldsymbol{\Sigma}) = \pi^N(\boldsymbol{B})\, \pi_u(\boldsymbol{\Sigma}) = C\, \pi_u(\boldsymbol{\Sigma}). \tag{5.8}$$

Note that this specification assumes that $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ are independent a priori, which is common when using noninformative improper priors (e.g. Jeffreys, 1961). Updating the prior in Equation (5.8) with the likelihood in Equation (5.5) results in a normal-inverse-Wishart posterior $N_{kp}\big(\boldsymbol{\beta}\big|\hat{\boldsymbol{\beta}}, \boldsymbol{\Sigma} \otimes (\boldsymbol{X}^T\boldsymbol{X})^{-1}\big) W_p^{-1}(\boldsymbol{\Sigma}|\boldsymbol{\Lambda} + \boldsymbol{S}, \nu + n - k)$ (a derivation is given in Appendix 5.A). From this it follows that the marginal unconstrained posterior distribution of $\boldsymbol{\Sigma}$ is given by

$$\pi_u(\boldsymbol{\Sigma}|\boldsymbol{Y}) = W_p^{-1}(\boldsymbol{\Sigma}|\boldsymbol{\Lambda} + \boldsymbol{S}, \nu + n - k). \tag{5.9}$$

Note that in order for this inverse-Wishart distribution to exist, the regressor matrix $\boldsymbol{X}$ must have full rank k and the sample size n must be greater than $p + k - \nu - 1$.

5.3 Bayes Factors for Testing Variances

5.3.1 The Bayes Factor

The Bayes factor for testing hypothesis H_t against a competing hypothesis $H_{t'}$ is defined as the ratio of the marginal likelihoods under the two hypotheses:

$$B_{tt'} = \frac{m_t(\boldsymbol{Y})}{m_{t'}(\boldsymbol{Y})}, \tag{5.10}$$

where $m_t(\boldsymbol{Y})$ is the marginal likelihood for observed data $\boldsymbol{Y}$ under hypothesis H_t with inequality constraints formulated according to Equation (5.1). The marginal likelihood is given by

$$m_t(\boldsymbol{Y}) = \int_{\Omega_t} \int_{\mathbb{R}^{kp}} f_t(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{B}, \boldsymbol{\Sigma}) \pi_t(\boldsymbol{B}, \boldsymbol{\Sigma})\, d\boldsymbol{B}\, d\boldsymbol{\Sigma}, \tag{5.11}$$

where $\Omega_t = \big\{\boldsymbol{\Sigma} : \boldsymbol{R}_t\boldsymbol{\sigma}^2 > \mathbf{0} \wedge \boldsymbol{\Sigma} > 0\big\}$ is the parameter space of $\boldsymbol{\Sigma}$ that is admissible under H_t and $\pi_t(\boldsymbol{B}, \boldsymbol{\Sigma})$ is the prior on $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ under H_t. Furthermore, $f_t(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{B}, \boldsymbol{\Sigma}) = f_u(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{B}, \boldsymbol{\Sigma})\, \mathbf{1}_{\Omega_t}(\boldsymbol{\Sigma})$, where $\mathbf{1}_{\Omega_t}(\boldsymbol{\Sigma})$ is an indicator function that equals 1 if $\boldsymbol{\Sigma} \in \Omega_t$ and 0 otherwise (i.e. the likelihood under H_t is a truncation of

the unconstrained likelihood in Equation (5.5)). The marginal likelihood can be interpreted as a weighted average likelihood in the parameter space that is admissible under H_t, where the weights are given by the prior. In this way the marginal likelihood quantifies how likely it is that the data were generated under H_t. The Bayes factor, as a ratio of marginal likelihoods, then tells us whether it is more likely that the data were generated under H_t or $H_{t'}$. If $B_{tt'} > 1$ ($B_{tt'} < 1$), then there is evidence in favor of H_t ($H_{t'}$). A Bayes factor of, say, 10 indicates that the evidence in favor of H_t is 10 times as strong as the evidence in favor of $H_{t'}$ (e.g. Jeffreys, 1961).

To facilitate the interpretation, in particular if there are more than two hypotheses under investigation, one can compute the posterior probabilities of the hypotheses using their marginal likelihoods and prior probabilities. The prior probabilities of the hypotheses, denoted by $P(H_1), \ldots, P(H_T)$, quantify how likely each hypothesis is before observing any data (the prior probabilities must sum to 1). If prior information about the hypotheses is weak, then it is customary to specify equal prior probabilities $P(H_1) = \cdots = P(H_T) = 1/T$. The posterior probabilities of the hypotheses are obtained by updating the prior probabilities using the marginal likelihoods:

$$P(H_t|\boldsymbol{Y}) = \frac{m_t(\boldsymbol{Y})P(H_t)}{\sum_{t'=1}^{T} m_{t'}(\boldsymbol{Y})P(H_{t'})}, \quad t = 1, \ldots, T. \tag{5.12}$$

The posterior probabilities $P(H_1|\boldsymbol{Y}), \ldots, P(H_T|\boldsymbol{Y})$ then quantify how likely each hypothesis is after observing the data (the posterior probabilities necessarily sum to 1).

5.3.2 Encompassing Prior Approach

From Equation (5.11) it can be seen that a prior $\pi_t(\boldsymbol{B}, \boldsymbol{\Sigma})$ on the coefficient and the covariance matrix needs to be specified in order to compute the marginal likelihood under hypothesis H_t. In this chapter we shall use the encompassing prior approach (Berger & Mortera, 1999; Klugkist et al., 2005) to specify priors under competing inequality constrained hypotheses on the variances in the multivariate normal linear model. The encompassing prior approach is a popular approach to computing Bayes factors between inequality constrained hypotheses. The starting point is the specification of an unconstrained (or encompassing) prior under the unconstrained hypothesis $H_u\colon \sigma_1^2, \ldots, \sigma_p^2 > 0$. Inequality constrained hypotheses are nested in the unconstrained hypothesis in the sense that $\Omega_t \subset \Omega_u = \{\boldsymbol{\Sigma} : \boldsymbol{\Sigma} > 0\}$. We may therefore specify priors under inequality constrained hypotheses as truncations of the unconstrained prior in the respective constrained subspaces. In this way, the problem of specifying a prior under each inequality constrained hypothesis simplifies to specifying one unconstrained prior.

The unconstrained prior $\pi_u(\boldsymbol{B}, \boldsymbol{\Sigma})$ on the coefficient and the covariance matrix was given in Equation (5.8). The prior under an inequality constrained hypothesis H_t is formulated as a truncation of the unconstrained prior in the admissible parameter space $\Omega_t \times \mathbb{R}^{kp}$:

$$\pi_t(\boldsymbol{B}, \boldsymbol{\Sigma}) = \frac{1}{P(\boldsymbol{\Sigma} \in \Omega_t|H_u)}\, \pi_u(\boldsymbol{B}, \boldsymbol{\Sigma})\, \mathbf{1}_{\Omega_t}(\boldsymbol{\Sigma}), \tag{5.13}$$

where

$$P(\boldsymbol{\Sigma} \in \Omega_t | H_u) = \int_{\Omega_t} \pi_u(\boldsymbol{\Sigma})\, d\boldsymbol{\Sigma} \tag{5.14}$$

is the prior probability that the inequality constraints on the variances on the main diagonal of $\boldsymbol{\Sigma}$ hold under H_u, with $\pi_u(\boldsymbol{\Sigma})$ as in Equation (5.7). The prior probability quantifies the complexity of an inequality constrained hypothesis relative to the unconstrained hypothesis, where a large prior probability indicates high complexity. In Equation (5.13) the inverse $1/P(\boldsymbol{\Sigma} \in \Omega_t | H_u)$ acts as a normalizing constant.

If prior information about the covariance matrix is weak, we recommend specifying the prior scale matrix of the unconstrained prior on $\boldsymbol{\Sigma}$ in Equation (5.7) as $\boldsymbol{\Lambda} = \text{diag}(\varepsilon, \ldots, \varepsilon)$, for a small positive ε. This choice can be considered objective for two reasons: First, under this setting all possible orderings of the variances $\left(\sigma_1^2 < \cdots < \sigma_p^2\right), \ldots, \left(\sigma_p^2 < \cdots < \sigma_1^2\right)$ are equally likely a priori. As was argued by Mulder, Hoijtink, and Klugkist (2010), such a specification is desirable when testing inequality constrained hypotheses and prior information is weak. Second, the contribution of the prior to the posterior scale matrix is negligible (in Equation (5.9) it can be seen that the posterior scale matrix is given by $\boldsymbol{\Lambda} + \boldsymbol{S} \approx \boldsymbol{S}$). As an uninformative choice for the prior degrees of freedom we recommend setting $\nu = p$, the smallest possible integer value. Then the prior contains minimal information in the sense that the contribution of the prior to the number of observations in the posterior is minimal.

Using the priors in Equations (5.8) and (5.13), it can be shown that the Bayes factor of an inequality constrained hypothesis H_t against the unconstrained hypothesis H_u is given by

$$B_{tu} = \frac{P(\boldsymbol{\Sigma} \in \Omega_t | \boldsymbol{Y}, H_u)}{P(\boldsymbol{\Sigma} \in \Omega_t | H_u)}, \tag{5.15}$$

where

$$P(\boldsymbol{\Sigma} \in \Omega_t | \boldsymbol{Y}, H_u) = \int_{\Omega_t} \pi_u(\boldsymbol{\Sigma} | \boldsymbol{Y})\, d\boldsymbol{\Sigma} \tag{5.16}$$

is the posterior probability that the inequality constraints on the variances on the main diagonal of $\boldsymbol{\Sigma}$ hold under H_u, with $\pi_u(\boldsymbol{\Sigma}|\boldsymbol{Y})$ as in Equation (5.9). A proof is given in Appendix 5.B. The posterior probability quantifies the fit of an inequality constrained hypothesis relative to the unconstrained hypothesis, where a large posterior probability indicates a good fit. From Equation (5.15) it can be seen that the proposed Bayes factor automatically functions as Occam's razor by taking the fit and the complexity of inequality constrained hypotheses into account. The Bayes factor between two inequality constrained hypotheses H_t and $H_{t'}$ is given by

$$B_{tt'} = \frac{B_{tu}}{B_{t'u}} = \frac{P(\boldsymbol{\Sigma} \in \Omega_t | \boldsymbol{Y}, H_u)/P(\boldsymbol{\Sigma} \in \Omega_t | H_u)}{P(\boldsymbol{\Sigma} \in \Omega_{t'} | \boldsymbol{Y}, H_u)/P(\boldsymbol{\Sigma} \in \Omega_{t'} | H_u)}. \tag{5.17}$$

Note that the posterior probability of an inequality constrained hypothesis H_t can be computed using the Bayes factors against the unconstrained hypothesis according to $P(H_t|\boldsymbol{Y}) = B_{tu}P(H_t) \big/ \left(\sum_{t'=1}^{T} B_{t'u}P(H_{t'})\right)$.

An important property of the proposed Bayes factor in Equation (5.17) is that it is consistent in the sense that the posterior probability of the true hypothesis tends to 1 as the sample size tends to infinity. This can be explained as follows: Suppose the true values of the variances lie in the admissible parameter space Ω_t of hypothesis

H_t. Then, as the sample size increases, the unconstrained posterior of $\boldsymbol{\Sigma}$ concentrates in Ω_t, so that the probability $P(\boldsymbol{\Sigma} \in \Omega_t | \boldsymbol{Y}, H_u)$ tends to 1. This implies that if $H_{t'}$ with admissible parameter space $\Omega_{t'}$ is a competing hypothesis that does not overlap with H_t (in the sense that $\Omega_t \cap \Omega_{t'} = \varnothing$), then $P(\boldsymbol{\Sigma} \in \Omega_{t'} | \boldsymbol{Y}, H_u)$ tends to 0. From Equation (5.17) it can be seen that the Bayes factor of H_t against $H_{t'}$ tends to infinity in this case (note that the prior probabilities $P(\boldsymbol{\Sigma} \in \Omega_t | H_u)$ and $P(\boldsymbol{\Sigma} \in \Omega_{t'} | H_u)$ are constants that do not depend on the data). As a result, the posterior probability of H_t tends to 1, which implies that the Bayes factor in Equation (5.17) is consistent. The same argument applies in the case where there are multiple competing hypotheses that do not overlap with H_t. To ensure that the Bayes factor behaves consistently it is advisable to specify the hypotheses such that they do not overlap and cover the entire unconstrained parameter space (in the sense that $\Omega_1 \cup \cdots \cup \Omega_T = \Omega_u$).

The probabilities in Equations (5.14) and (5.16) can be approximated numerically by drawing a large sample from the respective inverse-Wishart distribution and then computing the proportion of draws that fall in the admissible parameter space Ω_t. For example, to compute the posterior probability in Equation (5.16) we first sample $\boldsymbol{\Sigma}^{(m)} \sim W_p^{-1}(\boldsymbol{\Lambda} + \boldsymbol{S}, \nu + n - k)$, for $m = 1, \ldots, M$, where M is the number of draws and $\boldsymbol{\Sigma}^{(m)}$ is the mth draw. Then $P(\boldsymbol{\Sigma} \in \Omega_t | \boldsymbol{Y}, H_u) \approx \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}_{\Omega_t}\big(\boldsymbol{\Sigma}^{(m)}\big)$. Generating random draws from an inverse-Wishart distribution can for example be done in R (R Core Team, 2015) using the `riwish()` function from the MCMCpack package (A. D. Martin, Quinn, & Park, 2011).

5.4 Performance of the Bayes Factor

We conducted a simulation study investigating the performance of the Bayes factor in testing inequality constrained hypotheses on the variances in a 4×4 covariance matrix. Our focus of interest was the effect of the sample size on the evidence in favor of the true hypothesis as indicated by the Bayes factor and the posterior probability of the true hypothesis.

In our simulation design there were two factors. The first was the structure of the population covariance matrix. We used two population covariance matrices, one in which the variances followed an increasing order $\sigma_1^2 < \sigma_2^2 < \sigma_3^2 < \sigma_4^2$ and another in which the order of the first two variances was interchanged, that is, $\sigma_2^2 < \sigma_1^2 < \sigma_3^2 < \sigma_4^2$. To determine the population covariance matrices we used the separation strategy of Barnard, McCulloch, and Meng (2000), in which the covariance matrix is decomposed as $\boldsymbol{\Sigma} = \boldsymbol{DCD}$, where $\boldsymbol{D} = \text{diag}(\sigma_1, \ldots, \sigma_p)$ is a $p \times p$ matrix with the standard deviations on the main diagonal and $\boldsymbol{C} > 0$ is a $p \times p$ correlation matrix. For our simulation study we specified the population standard deviations as follows. First, we set $\sigma_1 = 1$ and $\sigma_4 = 1.58$. Research on heterogeneity of variances in the psychological sciences indicates that the resulting population variance ratio of $\sigma_4^2/\sigma_1^2 = 1.58^2/1^2 = 2.5$ can be considered a medium effect (Ruscio & Roche, 2012, Table 2). The intermediate population standard deviations were specified as $\sigma_2 = \sigma_4^{1/3} = 1.16$ and $\sigma_3 = \sigma_4^{2/3} = 1.36$, which resulted in constant ratios of adjacent standard deviations of $\sigma_2/\sigma_1 = \sigma_3/\sigma_2 = \sigma_4/\sigma_3 = 1.16$. We then defined two diagonal matrices $\boldsymbol{D}_1 = \text{diag}(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ and $\boldsymbol{D}_2 = \text{diag}(\sigma_2, \sigma_1, \sigma_3, \sigma_4)$, where the standard deviations on the main diagonal of $\boldsymbol{D}_1$ follow an increasing order, and in $\boldsymbol{D}_2$ the

first two standard deviations are interchanged. For the correlation matrix $\boldsymbol{C}$ we used a compound symmetry structure in which we set all off-diagonal elements equal to $\rho = 0.3$, which is considered a medium correlation in the social sciences (Cohen, 1992). Eventually, we computed the population covariance matrices as

$$\boldsymbol{\Sigma}_1 = \boldsymbol{D}_1\boldsymbol{C}\boldsymbol{D}_1 = \begin{bmatrix} 1.00 & 0.35 & 0.41 & 0.47 \\ 0.35 & 1.36 & 0.47 & 0.55 \\ 0.41 & 0.47 & 1.84 & 0.64 \\ 0.47 & 0.55 & 0.64 & 2.50 \end{bmatrix} \tag{5.18}$$

and

$$\boldsymbol{\Sigma}_2 = \boldsymbol{D}_2\boldsymbol{C}\boldsymbol{D}_2 = \begin{bmatrix} 1.36 & 0.35 & 0.47 & 0.55 \\ 0.35 & 1.00 & 0.41 & 0.47 \\ 0.47 & 0.41 & 1.84 & 0.64 \\ 0.55 & 0.47 & 0.64 & 2.50 \end{bmatrix}. \tag{5.19}$$

The second factor in our simulation study was the sample size $n \in \{5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000\}$.

In total there were 2 (covariance matrices) $\times$ 11 (sample sizes) = 22 conditions. In each condition we drew 1000 samples of size n from $N_4(\mathbf{0}, \boldsymbol{\Sigma}_1)$ and $N_4(\mathbf{0}, \boldsymbol{\Sigma}_2)$, respectively. In each sample we modeled the data as $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E} = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}^T \begin{bmatrix} \beta_{01} & \beta_{02} & \beta_{03} & \beta_{04} \end{bmatrix} + \boldsymbol{E}$ and tested the inequality constrained hypothesis $H_1 \colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2 < \sigma_4^2$ against its complement $H_2 \colon \neg H_1$. Specifically, we computed the logarithm of the Bayes factor (also referred to as the weight of evidence) in favor of the true hypothesis as well as the posterior probability of the true hypothesis assuming equal prior probabilities. Note that for data sampled from a multivariate normal distribution with covariance matrix $\boldsymbol{\Sigma}_1$ the true hypothesis is H_1, whereas for data sampled under $\boldsymbol{\Sigma}_2$ the complement H_2 is true. In computing the log Bayes factors and posterior probabilities we specified the hyperparameters of the unconstrained prior on the covariance matrix in Equation (5.7) as $\boldsymbol{\Lambda} = \text{diag}\left(10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}\right)$ and $\nu = p = 4$.

The top row of Figure 5.1 shows the simulation results for the covariance matrix $\boldsymbol{\Sigma}_1$, in which the variances were ordered as $\sigma_1^2 < \sigma_2^2 < \sigma_3^2 < \sigma_4^2$ (see Equation (5.18)). The top left plot shows the median of the 1000 log Bayes factors in favor of the true hypothesis H_1 (solid line) as well as the 2.5% and 97.5% quantiles (dashed lines) as a function of the sample size n. It can be seen that the evidence in favor of H_1 increased monotonically with the sample size. The variability of the evidence (as indicated by the quantiles) was largest for sample sizes up to 20. The log Bayes factor always selected the true hypothesis (i.e. it was always greater than 0) for sample sizes of 200 and above. The lines are discontinued because for large samples the posterior probability that the inequality constraints under H_1 do not hold was approximated as 0, so that the log Bayes factor was infinite (cf. Equation (5.17)). The top right plot in Figure 5.1 shows the median of the 1000 posterior probabilities of the true hypothesis H_1 as well as the 2.5% and 97.5% quantiles as a function of the sample size n. The results confirmed the findings from the log Bayes factor analysis: The evidence in favor of the true hypothesis H_1 increased monotonically with the sample size and there was considerable variability in the evidence for sample sizes up to 20. It can be seen that a sample size of 20 was sufficient to obtain a median posterior probability of H_1 above 0.8. The fact that the posterior probability approached 1

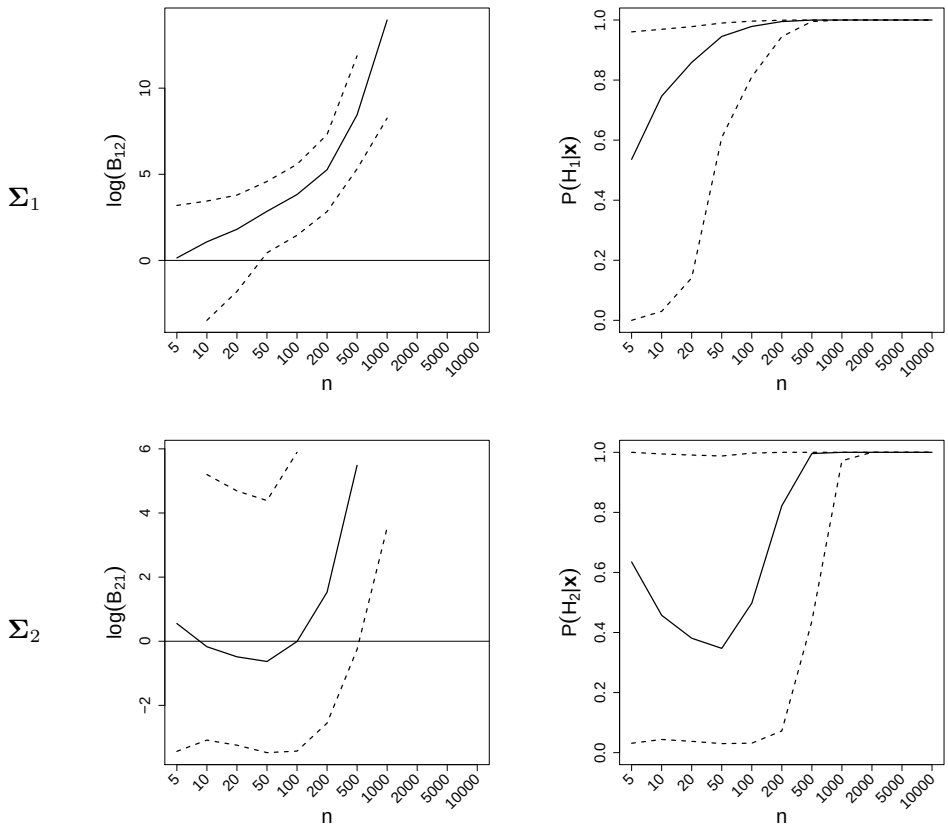


Figure 5.1: Results of a simulation study testing the inequality constrained hypothesis $H_1\colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2 < \sigma_4^2$ against its complement $H_2\colon \neg H_1$. The plots in the top (bottom) row show the results for a population covariance matrix $\mathbf{\Sigma}_1$ ($\mathbf{\Sigma}_2$) in which the variances are ordered as $\sigma_1^2 < \sigma_2^2 < \sigma_3^2 < \sigma_4^2$ ($\sigma_2^2 < \sigma_1^2 < \sigma_3^2 < \sigma_4^2$) (see Equations (5.18) and (5.19)). Note that for $\mathbf{\Sigma}_1$ ($\mathbf{\Sigma}_2$) the true hypothesis is H_1 (H_2). The plots in the left-hand column show the median log Bayes factor in favor of the true hypothesis (solid line) as well as the 2.5% and 97.5% quantiles (dashed lines) as a function of the sample size n. The lines are discontinued due to numerical reasons (see text). The plots in the right-hand column show the median posterior probability of the true hypothesis as well as the 2.5% and 97.5% quantiles.

as the sample size increased illustrates that the Bayes factor is consistent (as was explained in Section 5.3.2).

The bottom row of Figure 5.1 shows the simulation results for the covariance matrix $\mathbf{\Sigma}_2$, in which the variances were ordered as $\sigma_2^2 < \sigma_1^2 < \sigma_3^2 < \sigma_4^2$ (see Equation (5.19)). The bottom left plot shows the results for the log Bayes factor in favor of the true hypothesis H_2. The lines are discontinued for reasons analogous to those in the top left plot. The results differ from those shown in the top left plot in three important ways: First, larger samples were required to obtain decisive evidence in favor of H_2. Second, it can be seen that the median log Bayes factor first decreased slightly until a sample size of 50 was reached, after which it increased monotonically. Third, there was more variability in the evidence, which can be seen from larger differences between the 97.5% and 2.5% quantiles. All three patterns were caused by the fact the true order of the variances in the population ($\sigma_2^2 < \sigma_1^2 < \sigma_3^2 < \sigma_4^2$) differed from $H_1 \colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2 < \sigma_4^2$ only in the order of one pair of variances. As a result, it was difficult for the log Bayes factor to detect the true hypothesis if the sample size was small. Nevertheless, the log Bayes factor always selected the true hypothesis H_2 for sample sizes of 2000 and above. The same patterns were present in the results for the posterior probability of the true hypothesis shown in the bottom right plot. It can be seen that larger samples were necessary to obtain decisive evidence in favor of H_2, that the median posterior probability of H_2 first decreased, and that there was substantial variability for sample sizes up to 500. A sample size of 200 was required to obtain a median posterior probability of the true complement above 0.8. Again, we see consistent behavior in that the posterior probability of the true hypothesis approached 1 as the sample size increased.

5.5 Example Application: Reading Recognition in Children

We applied the Bayes factor to an empirical data set discussed in Vermunt and Magidson (2005). The outcome of interest in this data set was reading recognition in children as measured by the Reading Recognition subtest of the Peabody Individual Achievement Test (Dunn & Markwardt, 1970). The Reading Recognition subtest measures a child's word recognition and pronunciation ability. The subtest was administered four times at two-year intervals to a sample of 405 children. The data set contained the following background variables, which we used as covariates: the child's gender, the mother's age, the child's age, the child's cognitive stimulation at home, and the child's emotional support at home. Following Aunola et al. (2004), we hypothesized that the variability of reading recognition performances either increased or decreased over time. This corresponds to the two inequality constrained hypotheses $H_1 \colon \sigma_1^2 < \sigma_2^2 < \sigma_3^2 < \sigma_4^2$ and $H_2 \colon \sigma_4^2 < \sigma_3^2 < \sigma_2^2 < \sigma_1^2$, where σ_j^2 is the variance of the jth measurement. We tested these hypotheses against their complement $H_3 \colon \neg\,(H_1 \vee H_2)$ to take into account that the data might support neither H_1 nor H_2.

We fitted a multivariate normal linear model to the 233 complete cases with reading recognition as the 4-dimensional dependent variable and the background variables as independent variables (only main effects). The maximum likelihood estimate of

the covariance matrix was given by

$$\hat{\boldsymbol{\Sigma}} = \boldsymbol{S}/n = \begin{bmatrix} 0.50 & 0.38 & 0.38 & 0.39 \\ 0.38 & 0.84 & 0.74 & 0.77 \\ 0.38 & 0.74 & 1.06 & 0.92 \\ 0.39 & 0.77 & 0.92 & 1.28 \end{bmatrix}. \tag{5.20}$$

It can be seen that the point estimates of the variances follow an increasing order. Furthermore, note that the sample variance ratio of $\hat{\sigma}_4^2/\hat{\sigma}_1^2 = 1.28/0.50 = 2.57$ indicates a medium effect (similar to the effect size of $\sigma_4^2/\sigma_1^2 = 2.5$ used in the simulation study). To determine the evidence in the data in favor of H_1, H_2, and H_3, we computed the Bayes factors between the hypotheses and the posterior probabilities of the hypotheses. We specified the hyperparameters of the unconstrained prior on the covariance matrix as $\boldsymbol{\Lambda} = \text{diag}\left(10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}\right)$ and $\nu = p = 4$ (as we did in the simulation study). The resulting Bayes factors were given by $B_{12} = \infty$, $B_{13} = 1456.49$, and $B_{23} = 0.00$. The Bayes factors B_{12} and B_{23} were infinite and 0, respectively, because the posterior probability that the variances follow a decreasing order as stated in H_2 was approximated as 0. The results show that hypothesis H_1 stating an increase in variance received very strong support, while there was no support for either H_2 or H_3. This can also be seen from the posterior probabilities of the hypotheses, which were given by $P(H_1|\boldsymbol{Y}) = 1.00$, $P(H_2|\boldsymbol{Y}) = 0.00$, and $P(H_3|\boldsymbol{Y}) = 0.00$ (assuming equal prior probabilities). Considering the Bayes factors and posterior probabilities of the hypotheses, we ruled out H_2 and H_3 and concluded that the variances of reading recognition performances increased over time. Following Aunola et al. (2004), a possible explanation for such an increase in the variances is that children with high reading skills increased their reading recognition performances more over time than children with low reading skills.

5.6 Conclusion

In this chapter we developed a Bayes factor for testing inequality constrained hypotheses on the variances of dependent observations. We applied the encompassing prior approach, in which priors under inequality constrained hypotheses are formulated as truncations of the prior under the unconstrained hypothesis. This approach to prior specification has two main advantages: First, the problem of specifying a prior under every inequality constrained hypothesis to be tested simplifies to specifying one unconstrained prior. This is a useful property in practice, since in our experience researchers find it difficult to formulate subjective priors on (co)variances, in particular under inequality constrained hypotheses. We used a conjugate inverse-Wishart distribution with noninformative hyperparameters as the unconstrained prior on the covariance matrix. With this specification the Bayes factor is fully automatic in the sense that there is no need to specify a subjective prior under any hypothesis. The second advantage of the encompassing prior approach is that Bayes factors between inequality constrained hypotheses can be computed straightforwardly by sampling from the prior and the posterior distribution of the covariance matrix and computing the proportion of draws that satisfy the inequality constraints. Our Bayes factor is consistent in that it is guaranteed to select the true hypothesis as the sample size increases.

In our simulation study we considered a limited number of factors and conditions. Future research may examine the Bayes factor's behavior under different dimensions p, additional orderings and ratios of population variances, different magnitudes and structures of correlations, and additional hypotheses that are tested. Furthermore, it would be interesting to investigate relations between the multivariate normal linear model and other models of development over time. For example, in a random slope model the variance can either decrease over time, increase over time, or first decrease and then increase (e.g. Snijders & Bosker, 2012). The Bayes factor we presented in this chapter could be used to test such patterns in the variances of repeated measures. A natural extension of the testing problem we considered is to allow for equality constraints between the variances (e.g. $H\colon \sigma_1^2 = \sigma_2^2 < \sigma_3^2$). Here the challenge is to incorporate the equality constraints when computing the marginal likelihoods under competing hypotheses. The separation strategy of Barnard, McCulloch, and Meng (2000) could be a useful approach to modeling the covariances in the multivariate normal linear model when equality constraints are imposed on the variances. Based on our findings in this chapter we expect that the Bayes factor provides a solution to this intricate testing problem.

5.A Posterior Distribution of $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$

The joint unconstrained posterior distribution of the coefficient matrix $\boldsymbol{B}$ and the covariance matrix $\boldsymbol{\Sigma}$ is given by

$$\begin{aligned}
\pi_u(\boldsymbol{B}, \boldsymbol{\Sigma}|\boldsymbol{Y}) &\propto f_u(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{B}, \boldsymbol{\Sigma})\pi_u(\boldsymbol{B}, \boldsymbol{\Sigma}) \\
&\propto |\boldsymbol{\Sigma}|^{-n/2} \exp\left(-\tfrac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \left(\boldsymbol{\Sigma} \otimes (\boldsymbol{X}^T\boldsymbol{X})^{-1}\right)^{-1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right) \\
&\qquad \exp\left(-\tfrac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S})\right) |\boldsymbol{\Sigma}|^{-(\nu+p+1)/2} \exp\left(-\tfrac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda})\right) \\
&\propto \left|\boldsymbol{\Sigma} \otimes (\boldsymbol{X}^T\boldsymbol{X})^{-1}\right|^{-1/2} \exp\left(-\tfrac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \left(\boldsymbol{\Sigma} \otimes (\boldsymbol{X}^T\boldsymbol{X})^{-1}\right)^{-1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right) \\
&\qquad |\boldsymbol{\Sigma}|^{-(\nu+n-k+p+1)/2} \exp\left(-\tfrac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Lambda} + \boldsymbol{S})\right)\right),
\end{aligned} \tag{5.21}$$

which can be recognized as a normal-inverse-Wishart distribution $N_{kp}\big(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}, \boldsymbol{\Sigma} \otimes (\boldsymbol{X}^T\boldsymbol{X})^{-1}\big)W_p^{-1}(\boldsymbol{\Sigma}|\boldsymbol{\Lambda} + \boldsymbol{S}, \nu + n - k)$. From this it follows that the marginal unconstrained posterior distribution of $\boldsymbol{\Sigma}$ is given by $\pi_u(\boldsymbol{\Sigma}|\boldsymbol{Y}) = W_p^{-1}(\boldsymbol{\Sigma}|\boldsymbol{\Lambda} + \boldsymbol{S}, \nu + n - k)$. Note that the model assumes that $\operatorname{rank}(\boldsymbol{X}) = k$ and that $n > p + k - \nu - 1$.

5.B Bayes Factor of H_t Against H_u

The Bayes factor of an inequality constrained hypothesis H_t against the unconstrained hypothesis H_u is given by

$$\begin{aligned}
B_{tu} &= \frac{m_t(\boldsymbol{Y})}{m_u(\boldsymbol{Y})} = \frac{\int_{\Omega_t}\int_{\mathbb{R}^{kp}} f_t(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{B},\boldsymbol{\Sigma})\pi_t(\boldsymbol{B},\boldsymbol{\Sigma})\,d\boldsymbol{B}\,d\boldsymbol{\Sigma}}{\int_{\Omega_u}\int_{\mathbb{R}^{kp}} f_u(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{B},\boldsymbol{\Sigma})\pi_u(\boldsymbol{B},\boldsymbol{\Sigma})\,d\boldsymbol{B}\,d\boldsymbol{\Sigma}} \\
&= \frac{\int_{\Omega_t}\int_{\mathbb{R}^{kp}} f_u(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{B},\boldsymbol{\Sigma})P(\boldsymbol{\Sigma}\in\Omega_t|H_u)^{-1}\pi_u(\boldsymbol{B},\boldsymbol{\Sigma})\mathbf{1}_{\Omega_t}(\boldsymbol{\Sigma})\,d\boldsymbol{B}\,d\boldsymbol{\Sigma}}{\int_{\Omega_u}\int_{\mathbb{R}^{kp}} f_u(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{B},\boldsymbol{\Sigma})\pi_u(\boldsymbol{B},\boldsymbol{\Sigma})\,d\boldsymbol{B}\,d\boldsymbol{\Sigma}} \\
&= \frac{1}{P(\boldsymbol{\Sigma}\in\Omega_t|H_u)}\int_{\Omega_t}\int_{\mathbb{R}^{kp}} \frac{f_u(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{B},\boldsymbol{\Sigma})\pi_u(\boldsymbol{B},\boldsymbol{\Sigma})}{\int_{\Omega_u}\int_{\mathbb{R}^{kp}} f_u(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{B},\boldsymbol{\Sigma})\pi_u(\boldsymbol{B},\boldsymbol{\Sigma})\,d\boldsymbol{B}\,d\boldsymbol{\Sigma}}\,d\boldsymbol{B}\,d\boldsymbol{\Sigma} \\
&= \frac{1}{P(\boldsymbol{\Sigma}\in\Omega_t|H_u)}\int_{\Omega_t}\int_{\mathbb{R}^{kp}} \pi_u(\boldsymbol{B},\boldsymbol{\Sigma}|\boldsymbol{Y})\,d\boldsymbol{B}\,d\boldsymbol{\Sigma} \\
&= \frac{1}{P(\boldsymbol{\Sigma}\in\Omega_t|H_u)}\int_{\Omega_t} \pi_u(\boldsymbol{\Sigma}|\boldsymbol{Y})\,d\boldsymbol{\Sigma} = \frac{P(\boldsymbol{\Sigma}\in\Omega_t|\boldsymbol{Y},H_u)}{P(\boldsymbol{\Sigma}\in\Omega_t|H_u)},
\end{aligned} \tag{5.22}$$

where in the third line we may omit the indicator function because the integration region is already restricted to the constrained parameter space Ω_t. A derivation of the posterior distribution $\pi_u(\boldsymbol{B},\boldsymbol{\Sigma}|\boldsymbol{Y})$ is given in Appendix 5.A.

Chapter 6

Epilogue

In this dissertation we argued that variances are more than just nuisance parameters and that by disregarding the variability of observations one runs the risk of overlooking important information in the data. We provided numerous examples where specific relations between variances could be expected based on theoretical considerations or findings from previous research. For example, in the illustrative example in Section 2.6.1, research on gender differences suggested that boys are more variable in their intellectual abilities than girls. In the examples in Sections 3.3/3.6 and 5.5, theoretical considerations suggested that the variability of educational performances either increases or decreases over time. In Section 4.5.2 it was expected, based on previous research, that ADHD patients are more variable in their intellectual performances than unaffected controls and that Tourette's patients are either as variable as ADHD patients or unaffected controls. Such expectations were translated into hypotheses with equality and/or inequality constraints on the variances. Two competing hypotheses we frequently considered were the null hypothesis stating equality of variances and the complement of a set of hypotheses.

We used the Bayes factor to test equality and inequality constrained hypotheses on variances. Application of Bayes factors requires a prior distribution of the model parameters under every hypothesis to be tested. In this dissertation we developed Bayes factors based on automatic priors, which circumvent the need for the user to manually specify subjective priors. In Chapters 2, 3, and 4 we proposed Bayes factors for testing variances of independent observations. Here we used information from the sample data to specify proper priors in an automatic fashion. The priors were specified such that they contain minimal information, which avoids the Jeffreys–Lindley paradox (Jeffreys, 1961; Lindley, 1957) when testing hypotheses with equality and inequality constraints on the variances. The Bayes factor in Chapter 5 was based on a default unconstrained prior where the hyperparameters were specified such that the prior was uninformative when testing inequality constrained hypotheses on the variances of dependent observations.

Our Bayes factors have the following desirable properties when testing (in)equality constrained hypotheses on variances. First, the Bayes factors are able to quantify the evidence in the data in favor of the hypotheses under investigation. Second, it is straightforward to test multiple (non)nested hypotheses simultaneously. The

Bayes factors then indicate which hypothesis receives strongest support from the data. Third, results from simulation studies indicated that the Bayes factors are consistent in the sense that they select the true hypothesis as the sample size increases. In Chapter 5 consistency of the Bayes factor for testing variances of dependent observations was explained theoretically based on the behavior of the posterior distribution for large samples. Fourth, our Bayes factors function as Occam's razor by automatically taking the parsimony introduced by inequality constraints between variances into account. Note that this did not hold for the fractional Bayes factor in Chapters 2 and 4 if there was very strong evidence in favor of an inequality constrained hypothesis. Fifth, our Bayes factors are fully automatic, which means that users do not need to specify any prior distributions themselves.

The Bayes factors we proposed in this dissertation use normal distributions to model the data. The normal distribution, however, may not be a suitable model for data that depart in important aspects from normality. Such deviations from normality include outliers, skewness, and kurtosis. More research is needed to investigate how robust the proposed Bayes factors are to violations of normality. Moreover, further research is indicated to examine the Bayes factors' behavior under conditions different from those we considered in our simulation studies. Additional simulations might consider different numbers of populations (or dimensions in the case of dependent observations), different structures of population variances, and different effect sizes. Furthermore, in the case of dependent observations it would be interesting to investigate the behavior of the Bayes factor under different correlation matrices (different correlations in the compound symmetry structure and/or different correlation structures).

Our approach to testing inequality constrained hypotheses on the variances in the multivariate normal linear model in Chapter 5 can be extended in multiple ways. First, a natural extension is to enable both equality and inequality constraints between the variances. In such a testing problem the challenge is to include the equality constraints in the computation of the marginal likelihoods under competing hypotheses. The separation strategy of Barnard, McCulloch, and Meng (2000) might be a useful approach to achieve this. Here the covariance matrix is decomposed into standard deviations and correlations. Priors can then be formulated on the unique standard deviations using a balanced approach with equal priors on the standard deviations or a fractional approach by means of the (adjusted) fractional Bayes factor. As a second extension, one might consider including time-varying covariates in the multivariate normal linear model, which is relevant in repeated measurement studies (e.g. Mulder et al., 2009). A third extension would be testing variances in MAN(C)OVA models. These models commonly assume homoscedasticity, which means that each multivariate population has the same covariance matrix. Hypotheses can then be formulated on the common variances. The assumption of homoscedasticity can be relaxed by allowing different populations to have separate covariance matrices. In this case hypotheses can be formulated on variances within and across covariance matrices of different populations. A fourth potential extension would be testing (in)equality constrained hypotheses on variances in generalized multivariate regression models. One such model is the seemingly unrelated regressions model of Zellner (1962), which differs from the standard multivariate regression model in that each dependent variable may have a separate regressor matrix.

Further directions for future research in the area of testing hypotheses on variances are conceivable. First, one might consider using informative priors when testing variances. Here the difficulty is to elicit priors under competing hypotheses based on prior information about the variances that is available before observing the data. This is particularly challenging under equality and inequality constrained hypotheses on variances. Second, in normal linear regression one might consider modeling the mean as well as the variance as a function of predictors (e.g. Harvey, 1976; Polasek, Liu, & Jin, 1998). Different sets of predictors could be used for the mean and the variance, based on which variables researchers expect these to depend on. Hypotheses could then be formulated on the regression coefficients for the variance. For example, if a researcher expects males to be more variable than females, the corresponding hypothesis would be that the regression coefficient of the predictor "gender" for the variance is greater than 0 (if male is coded as 1). Third, variability could also be investigated in distributions other than the normal. For example, in generalized linear models the interest could focus not only on the mean but also on the dispersion parameter. Here the approach of Smyth (1989) might prove useful, where the mean as well as the dispersion is modeled as a function of (potentially different) predictors. Similar to the normal case above, researchers could then formulate hypotheses on the regression coefficients for the dispersion. A fourth direction would be testing hypotheses on variance components in random effects models. For example, Mulder and Fox (2013) proposed Bayes factors for testing hypotheses on variance components in random intercept models. This work could be extended to more general random effects models such as random slopes models. Based on the findings in this dissertation I expect that the Bayes factor provides a solution to the testing problems outlined above.

A software application for computing the adjusted fractional Bayes factor was presented in Chapter 3. In the future it would be valuable to also make software available for computing the fractional Bayes factor and the balanced Bayes factor from Chapters 2 and 4, as well as the Bayes factor for testing variances of dependent observations from Chapter 5. Here it would be useful to combine all four Bayes factors in one user-friendly package in R (R Core Team, 2015). Major parts of the computer code for such a package are already available since all computations in this dissertation were carried out in R. Currently, many separate software packages exist for testing (in)equality constrained hypotheses using Bayes factors. Some examples include BIEMS for testing (in)equality constrained hypotheses on means and regression coefficients in the multivariate normal linear model (Mulder et al., 2012), BIG for computing the approximate Bayes factor presented in Gu et al. (2014), and BOCOR for testing inequality constrained hypotheses on correlations (Mulder, 2016). In the long run it would be useful to combine all these different approaches in one software program for testing (in)equality constrained hypotheses. A challenge would be that the existing packages are written in different programming languages. Nevertheless, such a unifying program would be a valuable resource for applied researchers to test specific expectations about various parameters.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csáki (Eds.), *2nd international symposium on information theory* (pp. 267–281). Budapest, Hungary: Akadémiai Kiadó.

Anraku, K. (1999). An information criterion for parameters under a simple order restriction. *Biometrika*, *86*(1), 141–152.

Arden, R., & Plomin, R. (2006). Sex differences in variance of intelligence across childhood. *Personality and Individual Differences*, *41*(1), 39–48.

Aunola, K., Leskinen, E., Lerkkanen, M.-K., & Nurmi, J.-E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology*, *96*(4), 699–713.

Barnard, J., McCulloch, R., & Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, *10*(4), 1281–1311.

Bartlett, M. S. (1957). A comment on D. V. Lindley's statistical paradox. *Biometrika*, *44*(3–4), 533–534.

Berger, J. O. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, *1*(3), 385–402.

Berger, J. O., & Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, *2*(3), 317–335.

Berger, J. O., & Mortera, J. (1999). Default Bayes factors for nonnested hypothesis testing. *Journal of the American Statistical Association*, *94*(446), 542–554.

Berger, J. O., & Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, *91*(433), 109–122.

Berger, J. O., & Pericchi, L. R. (2001). Objective Bayesian methods for model selection: Introduction and comparison. In P. Lahiri (Ed.), *Model selection* (pp. 135–207). Beachwood, OH: Institute of Mathematical Statistics.

Berger, J. O., & Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of P values and evidence. *Journal of the American Statistical Association*, *82*(397), 112–122.

Böing-Messing, F., & Mulder, J. (2016). Automatic Bayes factors for testing variances of two independent normal distributions. *Journal of Mathematical Psychology*, *72*, 158–170.

Böing-Messing, F., van Assen, M. A. L. M., Hofman, A. D., Hoijtink, H., & Mulder, J. (2017). Bayesian evaluation of constrained hypotheses on variances of multiple independent groups. *Psychological Methods*, *22*(2), 262–287.

Borkenau, P., Hřebíčková, M., Kuppens, P., Realo, A., & Allik, J. (2013). Sex differences in variability in personality: A study in four samples. *Journal of Personality*, *81*(1), 49–60.

Bryk, A. S., & Raudenbush, S. W. (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin*, *104*(3), 396–404.

Carroll, R. J. (2003). Variances are not always nuisance parameters. *Biometrics*, *59*(2), 211–220.

Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2015). *shiny: Web application framework for R.* Retrieved from `http://CRAN.R-project.org/package=shiny`

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Conigliani, C., & O'Hagan, A. (2000). Sensitivity of the fractional Bayes factor to prior distributions. *Canadian Journal of Statistics*, *28*(2), 343–352.

De Santis, F., & Spezzaferri, F. (2001). Consistent fractional Bayes factor for nested normal linear models. *Journal of Statistical Planning and Inference*, *97*(2), 305–321.

Dunn, L. M., & Markwardt, F. C. (1970). *Peabody Individual Achievement Test.* Circle Pines, MN: American Guidance Service.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.

Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research*, *62*(1), 61–84.

Gastwirth, J. L., Gel, Y. R., & Miao, W. (2009). The impact of Levene's test of equality of variances on statistical theory and practice. *Statistical Science*, *24*(3), 343–360.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016.

Gilks, W. R. (1995). Discussion of O'Hagan. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 118–120.

Grissom, R. J. (2000). Heterogeneity of variance in clinical data. *Journal of Consulting and Clinical Psychology*, *68*(1), 155–165.

Gu, X., Mulder, J., Deković, M., & Hoijtink, H. (2014). Bayesian evaluation of inequality constrained hypotheses. *Psychological Methods*, *19*(4), 511–527.

Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica*, *44*(3), 461–465.

Held, L. (2010). A nomogram for P values. *BMC Medical Research Methodology*, *10*(1), 21.

Hoijtink, H. (2011). *Informative hypotheses: Theory and practice for behavioral and social scientists.* Boca Raton, FL: Chapman & Hall/CRC.

Hubbard, R., & Armstrong, J. S. (2006). Why we don't really know what statistical significance means: Implications for educators. *Journal of Marketing Education*, *28*(2), 114–120.

Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *Journal of Gerontology: Psychological Sciences*, *57B*(2), 101–115.

Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability. *Mathematical Proceedings of the Cambridge Philosophical Society*, *31*(2), 203–222.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, England: Oxford University Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Kim, J., & Seltzer, M. (2011). Examining heterogeneity in residual variance to detect differential response to treatments. *Psychological Methods*, *16*(2), 192–208.

Klinkenberg, S., Straatemeier, M., & van der Maas, H. (2011). Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, *57*(2), 1813–1824.

Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality constrained analysis of variance: A Bayesian approach. *Psychological Methods*, *10*(4), 477–493.

Kofler, M. J., Rapport, M. D., Sarver, D. E., Raiker, J. S., Orban, S. A., Friedman, L. M., & Kolomeyer, E. G. (2013). Reaction time variability in ADHD: A meta-analytic review of 319 studies. *Clinical Psychology Review*, *33*(6), 795–811.

Lehre, A.-C., Lehre, K. P., Laake, P., & Danbolt, N. C. (2009). Greater intrasex phenotype variability in males than in females is a fundamental aspect of the gender differences in humans. *Developmental Psychobiology*, *51*(2), 198–206.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, & H. B. Mann (Eds.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (pp. 278–292). Stanford, CA: Stanford University Press.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, *103*(481), 410–423.

Lindley, D. V. (1957). A statistical paradox. *Biometrika*, *44*(1–2), 187–192.

Lucas, J. W. (2003). Status processes and the institutionalization of women as leaders. *American Sociological Review*, *68*(3), 464–480.

Maris, G., & van der Maas, H. (2012). Speed-accuracy response models: Scoring rules based on response time and accuracy. *Psychometrika*, *77*(4), 615–633.

Martin, A. D., Quinn, K. M., & Park, J. H. (2011). MCMCpack: Markov chain Monte Carlo in R. *Journal of Statistical Software*, *42*(9), 1–21.

Martin, N. A. J., Lundy, J. B., & Rickard, R. F. (2014). Lack of precision of burn surface area calculation by UK Armed Forces medical personnel. *Burns*, *40*(2), 246–250.

Mulder, J. (2014a). Bayes factors for testing inequality constrained hypotheses: Issues with prior specification. *British Journal of Mathematical and Statistical Psychology*, *67*(1), 153–171.

Mulder, J. (2014b). Prior adjusted default Bayes factors for testing (in)equality constrained hypotheses. *Computational Statistics & Data Analysis*, *71*, 448–463.

Mulder, J. (2016). Bayes factors for testing order-constrained hypotheses on correlations. *Journal of Mathematical Psychology*, *72*, 104–115.

Mulder, J., & Fox, J.-P. (2013). Bayesian tests on components of the compound symmetry covariance matrix. *Statistics and Computing*, *23*(1), 109–122.

Mulder, J., Hoijtink, H., & de Leeuw, C. (2012). BIEMS: A Fortran 90 program for calculating Bayes factors for inequality and equality constrained models. *Journal of Statistical Software*, *46*(2), 1–39.

Mulder, J., Hoijtink, H., & Klugkist, I. (2010). Equality and inequality constrained multivariate linear models: Objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference*, *140*(4), 887–906.

Mulder, J., Klugkist, I., van de Schoot, R., Meeus, W. H., Selfhout, M., & Hoijtink, H. (2009). Bayesian model selection of informative hypotheses for repeated measurements. *Journal of Mathematical Psychology*, *53*(6), 530–546.

Mulder, J., & Wagenmakers, E.-J. (in preparation). Editors' introduction to the special issue 'Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments'. *Journal of Mathematical Psychology*.

O'Hagan, A. (1995). Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 99–138.

O'Hagan, A. (1997). Properties of intrinsic and fractional Bayes factors. *Test*, *6*(1), 101–118.

Pérez, J. M., & Berger, J. O. (2002). Expected-posterior prior distributions for model selection. *Biometrika*, *89*(3), 491–511.

Polasek, W., Liu, S., & Jin, S. (1998). Heteroskedastic linear regression models: A Bayesian analysis. In I. Balderjahn, R. Mathar, & M. Schader (Eds.), *Classification, data analysis, and data highways* (pp. 182–191). Berlin, Germany: Springer.

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from `https://www.R-project.org/`

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237.

Ruscio, J., & Roche, B. (2012). Variance heterogeneity in published psychological research: A review and a new index. *Methodology*, *8*(1), 1–11.

Russell, V. A., Oades, R. D., Tannock, R., Killeen, P. R., Auerbach, J. G., Johansen, E. B., & Sagvolden, T. (2006). Response variability in Attention-Deficit/Hyperactivity Disorder: a neuronal and glial energetics hypothesis. *Behavioral and Brain Functions*, *2*(1), 1–25.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Sellke, T., Bayarri, M. J., & Berger, J. O. (2001). Calibration of p values for testing precise null hypotheses. *The American Statistician*, *55*(1), 62–71.

Silverstein, S. M., Como, P. G., Palumbo, D. R., West, L. L., & Osborn, L. M. (1995). Multiple sources of attentional dysfunction in adults with Tourette's syndrome: Comparison with attention deficit-hyperactivity disorder. *Neuropsychology*, *9*(2), 157–164.

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood

regression with beta-distributed dependent variables. *Psychological Methods*, *11*(1), 54–71.

Smyth, G. K. (1989). Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society. Series B (Methodological)*, *51*(1), 47–60.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London, England: Sage.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639.

Spiegelhalter, D. J., & Smith, A. F. M. (1982). Bayes factors for linear and log-linear models with vague prior information. *Journal of the Royal Statistical Society. Series B (Methodological)*, *44*(3), 377–387.

Straatemeier, M. (2014). *Math Garden: A new educational and scientific instrument* (Doctoral dissertation, University of Amsterdam, Amsterdam, the Netherlands). Retrieved from `http://dare.uva.nl/record/1/417091`

Van Mechelen, I. (2009). A royal road to understanding the mechanisms underlying person-in-context behavior. *Journal of Research in Personality*, *43*(2), 179–186.

Verhagen, A. J., & Fox, J. P. (2013). Bayesian tests of measurement invariance. *British Journal of Mathematical and Statistical Psychology*, *66*(3), 383–401.

Vermunt, J. K., & Magidson, J. (2005). Structural equation modeling: Mixture models. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1922–1927). Chichester, England: Wiley.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

Weerahandi, S. (1995). ANOVA under unequal error variances. *Biometrics*, *51*(2), 589–599.

Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, *19*(6), 1057–1064.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, *57*, 348–368.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In P. K. Goel & A. Zellner (Eds.), *Bayesian inference and decision techniques: Essays in honor of Bruno de Finetti* (pp. 233–243). Amsterdam, The Netherlands: Elsevier.

Summary

Statistical models commonly consist of two types of parameters. On the one hand, there are mean parameters that describe the average in a population of subjects (e.g., students) on a measure of interest (e.g., intelligence). Suppose the average intelligence quotient (IQ) in some population of students is 105. Not all students have an IQ of exactly 105. Instead, the IQ varies from student to student; some students have an IQ above 105, others below 105. In statistical models, this variability is captured by variance parameters. A small variance means that the members of a population have similar values close to the average on the measure of interest, whereas a large variance indicates that the values deviate a lot from the population's average. Note that mean as well as variance parameters are unknown quantities. It is the objective of statistical analyses to provide information on these parameters based on data coming from subjects in the population under study.

When statistical models are applied in practice, researchers are commonly interested in the mean parameters describing the average in the population under study. For example, one might be interested in the average intelligence in a population of students. The information that is expressed by variance parameters is often disregarded in practice. In fact, some of the most common statistical models (e.g., ANOVA) assume homogeneity of variances. This assumption states that the variances of different populations (e.g., male and female students) are equal. However, psychological research has shown that the assumption of homogeneity of variances often is not tenable and that frequently there is evidence that variances are not equal. For example, psychological studies have found males to be more variable than females on a variety of measures (e.g., intelligence).

Often there are reasons to expect that certain populations of subjects are less variable or more variable than other populations. Alternatively, one might expect two or more populations two be equally variable. Such expectations can be expressed as hypotheses with equality and inequality constraints on the variances of the populations. For example, based on theoretical considerations and/or previous research, one might expect male students to be more variable than female students on a particular outcome measure (e.g., intelligence). This expectation can be formulated as an inequality constrained hypothesis on the variances of the two populations of students: "The variance of male students is greater than the variance of female students." Here the inequality constraint is expressed by the phrase 'is greater than.' There are two potential competing hypotheses in this example: "the variance of male students is less than the variance of female students" and "the variances of male and female students are equal." Equality and inequality constrained hypotheses can also be formulated

on the variances of more than two populations. For example, when investigating the variability of students' academic performances, one might hypothesize that "the variance in grade 1 is less than the variance in grade 2, which is less than the variance in grade 3." In other words, one might expect that the variance increases across grades. When there are more than two populations under investigation, we may also combine equality and inequality constraints (e.g., "variance 1 and variance 2 are equal and variance 3 is greater than variances 1 and 2").

To test their hypotheses, researchers collect data on subjects from the populations under study. The question is then which hypothesis is most supported by the data. In this dissertation we use the Bayes factor to test (in)equality constrained hypotheses on variances and quantify the support in the data in favor of competing hypotheses. The Bayes factor is a hypothesis testing approach from an area called Bayesian statistics. The defining characteristic of Bayesian statistics is that there are two sources of information about model parameters such as variances: the sample data and the prior distribution (called prior in short). The prior is a probability distribution describing what is known about the model parameters before collecting any data (e.g., from previous research). In other words, the prior describes which values are likely for the parameters before we can look at the data. In Bayesian hypothesis testing by means of the Bayes factor, a prior needs to be specified under each hypothesis to be tested. Thus, for the example of comparing the variances of male and female students, a prior needs to be specified under each of the two inequality constrained hypotheses and under the equality constrained hypothesis. In practice, however, prior information about the magnitude of variances is often unavailable or researchers would like to refrain from adding prior information to the analysis (e.g., to "let the data speak for themselves"). For such situations statisticians devised so-called default or automatic Bayes factors. In these methods, the priors are specified in an automatic fashion without the user needing to provide any prior information. Often, information from the sample data is used to automatically specify the priors.

In this dissertation, we develop automatic Bayes factors for testing equality and inequality constrained hypotheses on variances. In Chapter 2, we consider the problem of testing (in)equality constrained hypotheses on the variances of two independent populations. Populations are independent if there is no relation between subjects from different populations. Throughout this dissertation we assume that the outcome of interest (e.g., intelligence) is normally distributed in the populations under study. In Chapter 2 we are concerned with testing the following three hypotheses: "the variance of population 1 is less than the variance of population 2," "the variance of population 1 is greater than the variance of population 2," and "the variances of populations 1 and 2 are equal." We present three automatic Bayes factors for this multiple hypothesis test. The first is the fractional Bayes factor (FBF) introduced by O'Hagan (1995). The FBF is a general approach to Bayesian hypothesis testing and model selection that implicitly specifies priors in an automatic fashion using information from the sample data. In this dissertation we apply the FBF for the first time to the problem of testing (in)equality constrained hypotheses on variances. It is shown that the FBF may not properly take the parsimony of inequality constrained hypotheses into account. Note that inequality constrained hypotheses are more parsimonious than unconstrained hypotheses because the former allow fewer values of the variances. For example, the inequality constrained hypothesis "the variance of male students is greater than

the variance of female students" excludes the possibility that the variance of male students is less than the variance of female students. The FBF does not properly take the parsimony of inequality constrained hypotheses into account because the implicit automatic prior favors those inequality constraints that are likely according to the data. The fact that the FBF may not incorporate the parsimony due to inequality constraints is a problem because in hypothesis testing one should accept the most parsimonious hypothesis that is in line with the data (this principle is also referred to as Occam's razor).

As a solution to the problem, we propose two automatic Bayes factors that take the parsimony of inequality constrained hypotheses into account when testing two variances. The first is a balanced Bayes factor (BBF), which is based on identical priors on the two variances. The priors are specified in an automatic fashion using information from the sample data. An important characteristic of the BBF is that the two possible orderings of the two variances (variance 1 is less than variance 2, variance 1 is greater than variance 2) are equally likely according to the prior. This ensures that the BBF properly incorporates the parsimony of inequality constrained hypotheses. However, the fact that the same prior is used for the two variances results in shrinkage in the sense that this implies evidence for the hypothesis stating equality of variances. The second solution we present is an adjusted fractional Bayes factor (aFBF). In this approach we adjust the FBF such that the two possible orderings of the two variances are equally likely a priori (as in the BBF). Contrary to the BBF, the aFBF is not affected by shrinkage. In a simulation study the FBF, BBF, and aFBF showed similar behavior, with the aFBF providing slightly stronger evidence in favor of a true inequality constrained hypothesis than the FBF and BBF. Note that a hypothesis is true if it correctly describes the relation between the variances of the populations. We demonstrate the practical utility of the three Bayes factors in two real data examples dealing with the variability of intelligence in children and the precision of burn wound assessments.

In Chapter 3, we extend the aFBF to the problem of testing (in)equality constrained hypotheses on the variances of more than two populations. We first apply the standard FBF to the special case of testing an inequality constrained hypothesis on the variances of three populations. This test reveals that the FBF may not properly incorporate the parsimony of inequality constrained hypotheses in the case of more than two populations either. As with the FBF for testing two variances in Chapter 2, this happens because the implicit automatic prior in the FBF approach favors those inequality constraints that are likely according to the data. Next, we develop the aFBF for the general case of testing hypotheses with equality and inequality constraints on the variances of more than two populations. An application of this extended aFBF to the test of an inequality constrained hypothesis on three variances shows that it takes the parsimony introduced by the inequality constraints into account. As for the case of testing two variances in Chapter 2, this is due to the fact that in the aFBF approach all possible orderings of the variances are equally likely a priori. The results of a simulation study indicate that the aFBF is large sample consistent, which means that it is able to detect the true hypothesis if the sample is large enough (i.e., if we have enough data). An application of the aFBF to real data from the Math Garden online learning environment (https://www.mathsgarden.com/) revealed strong evidence that the variability of students' multiplication and division

abilities in elementary school increases across grades. We conclude this chapter with a presentation of a user-friendly software application for computing the aFBF. The only input that is required from the user is the data and the hypotheses to be tested.

In Chapter 4, we extend the FBF and BBF from Chapter 2 to the case of testing (in)equality constrained hypotheses on more than two variances. As the BBF for testing two variances in Chapter 2, the extended BBF is based on identical priors for all variances. We present a detailed comparison of the performance of the extended FBF and BBF as well as the extended aFBF from Chapter 3. First, we illustrate that the BBF and aFBF incorporate the parsimony of inequality constrained hypotheses when testing more than two variances, whereas the FBF may not do so. Next, we examine whether the three Bayes factors are information consistent when testing inequality constrained hypotheses. Loosely speaking, a Bayes factor is information consistent in this setting if the evidence in favor of an inequality constrained hypothesis keeps increasing as the inequality effect in the data increases (while keeping the sample size fixed). Numerical analyses indicated that the FBF and aFBF are information consistent, whereas the BBF showed information inconsistent behavior. The third property we investigate is large sample consistency. The results of a simulation study indicated that all three Bayes factors are large sample consistent. The BBF provided stronger evidence in favor of a true equality constrained hypothesis than the FBF and aFBF. This is due to the identical priors on the variances in the BBF approach, which imply evidence in favor of equality of variances. The FBF and aFBF yielded stronger evidence in favor of a true inequality constrained hypothesis than the BBF. We illustrate the use of the three Bayes factors by applying them to hypothetical data from an experimental study and real data from two studies investigating attentional performances of Tourette's and ADHD patients and the influence of group leaders, respectively.

In Chapter 5, we consider the problem of testing inequality constrained hypotheses on the variances of dependent populations. Populations are dependent if there is a relationship between subjects from different populations or if the same subjects are surveyed multiple times. In this chapter we focus on inequality constrained hypotheses and exclude equality constraints for reasons of technical complexity. We use the so-called encompassing prior approach to specify priors under competing inequality constrained hypotheses. The first step in this approach is to define a prior under the unconstrained hypothesis that does not impose any constraints on the variances. Subsequently, the prior under an inequality constrained hypothesis is defined as a truncation of this unconstrained prior in the region that is admissible under the respective hypothesis. For example, consider the inequality constrained hypothesis "the variance of population 1 is less than the variance of population 2." We first specify a prior on the two variances that does not involve any constraints. Note that under this unconstrained prior the variances may take on any possible combination of values. That is, under the unconstrained prior variance 1 may be less than or greater than variance 2. The prior under the inequality constrained hypothesis is then proportional to the unconstrained prior in the region where variance 1 is less than variance 2 and it is 0 where variance 1 is greater than variance 2. With this approach to specifying priors, Bayes factors between competing inequality constrained hypotheses can be computed using the probabilities that the inequality constraints hold before and after observing the data. An important property of our Bayes factor based on

the encompassing prior approach is that it is large sample consistent. Furthermore, it is fully automatic in the sense that the user does not need to manually specify any prior. We conclude this chapter with an application of our Bayes factor to real data from a study investigating reading ability in children, which was measured four times in the same children at two-year intervals. Our analysis revealed strong evidence that the variance of children's reading ability increases over time.

Acknowledgements

Completing this dissertation would not have been possible for me without the outstanding support of my supervisors, colleagues, family, and friends. First of all, I would like to thank my daily supervisor, Joris Mulder. Joris, you have truly been a magnificent supervisor for me throughout my PhD project. You did a fantastic job in supporting me to make this dissertation a success. In the beginning we had some difficulties. The project was challenging for me and we struggled to find a suitable topic for the first paper. But thanks to your tight supervision, your detailed and ever constructive feedback, your patience, and your encouragement we were able to overcome these difficulties and all the challenges we encountered while working on the subsequent papers. I really think that we worked great as a team, and that it is due to this teamwork that we managed to produce four papers that I am actually proud of (if I am allowed to say so at this point). Next, I would like to thank my second supervisor, Marcel van Assen. Marcel, I would like to thank you for your marvelous supervision while Joris was on sabbatical in the USA and for your superb feedback on all four papers in this dissertation. Your contributions were very important for the quality of the papers and for getting Chapters 2 and 3 published in two renowned journals. I have always very much enjoyed working together with you. Thank you Joris and Marcel for this great time!

Of course, my supervisors were not the only ones to support me during my time here in Tilburg. In fact, I would like to thank the entire Department of Methodology and Statistics (current and former colleagues) for supporting me in various ways and for making my time in the department so valuable. Thanks to you, there was not a single day in five years where I did not enjoy going to work. Even on stressful days you were always able to put a smile on my face. I would like to thank the head of our department, Jeroen Vermunt, for his statistical and managerial support. Jeroen, you were always available when I had questions relating to statistics or my position as a PhD student in our department. In particular, I would like to thank you for offering me the position as chief statistical consultant for the Elisabeth-TweeSteden hospital in Tilburg. I learned so much in this job not only about statistics, but also about solving practical problems in close collaboration with professionals from various areas. I think I could develop myself a lot in this job. Speaking about my consulting job for the hospital, credit is due to Jeroen, Marcel, and Joris for their constructive help when I encountered a problem which I was not sure how to solve properly. Next, I would like to thank the members of the Bayes lab group for their constructive and helpful feedback on my draft papers. I honestly think that the papers in this dissertation benefited greatly from your contributions. Thanks also to

everyone involved in teaching Introductory Statistics for Liberal Arts and Sciences and Academic Skills, in particular Luc van Baest, Jeroen Stekelenburg, Mirjam Keetels, and Wilco Emons. Thanks to you (and the nice students) I always enjoyed teaching at Tilburg University. Last but not least, I would like to thank Marieke Timmermans, Anne-Marie van der Heijden, and everyone else who was so kind to help me out when I had questions about statistics, my dissertation, teaching, or my position as a PhD student. Thank you all for making my job at Tilburg University so exceptional!

A special thanks also goes to my two amazing paranymphs, Robert Hillen and Leonie Vogelsmeier. Robert, it was an honor for me to be your paranymph and to help making the day of your defense a day you will never forget. I am truly happy that now you support me in this final stage of my PhD. It is astonishing how quickly the four years we spent together in the department passed by. I think it is safe to say that we had a great time here in Tilburg. Leonie, we got to know each other at a rather late stage of my PhD. You started working in our department about a year ago and we immediately got along very well. Just like me, you had a lot of questions in the beginning of your PhD and I was happy to help you out as far as I could. We had a great time in our beautiful office garden and now I am very happy that you support me as a paranymph for my defense. Robert and Leonie, I am glad and thankful that you two are behind me!

Next to my colleagues, I was lucky to have my family and friends support me throughout my time as a PhD student. First of all, I would like to thank my family. You were always there for me and I knew that I could always count on you. I think your unconditional support is one of the major reasons why I always felt confident and never had thoughts of giving up, even in difficult times. I love you all. Next to my family, I am lucky to have a large group of awesome friends both in the Netherlands and in Germany. In the last five years I had amazing experiences with you guys in Tilburg, Bocholt, Düsseldorf, and countless other cities. As with my family, I drew a lot of strength and power from the fact that I could always rely on you. Thank you for bringing so much happiness to my life and for always being there for me!

Tilburg, September 6, 2017 *Florian Böing-Messing*

There are often reasons to expect certain relations between the variances of multiple populations. For example, in an educational study one might expect that the variance of students' performances increases or decreases across grades. Alternatively, it might be expected that the variance is constant across grades. Such expectations can be formulated as equality and inequality constrained hypotheses on the variances of the students' performances. In this dissertation we develop automatic (or default) Bayes factors for testing such hypotheses. The methods we propose are based on default priors that are specified in an automatic fashion using information from the sample data. Hence, there is no need for the user to manually specify priors under competing (in)equality constrained hypotheses, which is a difficult task in practice. All the user needs to provide is the data and the hypotheses. Our Bayes factors then indicate to what degree the hypotheses are supported by the data and, in particular, which hypothesis receives strongest support.